From the Department of Medical Epidemiology and Biostatistics
Karolinska Institutet, Stockholm, Sweden

A Short Story of the Long Telomere

Evidence in Aging Epidemiology

Yiqiang Zhan

Stockholm 2018

Published by Karolinska Institutet.
Printed by E-Print AB 2018
Produced by LaTeX with template downloaded from https://www.latextemplates.com, the-legrand-orange-book(Version 2.3), CC BY-NC-SA 3.0.

ISBN 978-91-7676-981-2

A Short Story of the Long Telomere: Evidence in Aging Epidemiology

Thesis For Doctoral Degree (Ph.D.)

By

Yiqiang Zhan

Time and Location: kl 09:00, April 27, 2018, in the lecture hall Petrén, Nobels väg 12B, Karolinska Institutet, Solna

Principle Supervisor
Docent Sara Hägg
Department of Medical Epidemiology
and Biostatistics
Karolinska Institutet

Co-Supervisors
Professor Nancy L. Pedersen
Department of Medical Epidemiology
and Biostatistics
Karolinska Institutet

Professor Chandra A. Reynolds
Department of Psychology
University of California, Riverside

Opponent
Reader Tom Gaunt
MRC Integrative Epidemiology Unit
University of Bristol

Examination Board
Docent Neda Agahi
Aging Research Center
Karolinska Institutet

Professor Kamila Czene
Department of Medical Epidemiology
and Biostatistics
Karolinska Institutet

Docent Marita Cohn
Department of Biology
Lund University

To My Family

ABSTRACT

Telomeres are the sequences of nucleotides at the end of chromosomes. Each time a cell divides, telomeres become shorter. The length of telomeres can be replenished by an enzyme, telomerase. Telomere shortening is hypothesized as the biological origin of aging. Indeed, telomeres are shorter in people with various diseases than others. Whether these observed associations are causal or due to other factors that could be the common causes of both telomere shortening and diseases are largely unknown. In this thesis, we aimed to disentangle the relationship between telomere length and several aging-associated diseases and traits to enhance our understanding of the biology underpinning disease pathogenesis.

In Study I, we performed a Mendelian randomization (MR) study examining the association between telomere length and Alzheimer's disease, using genome-wide association study (GWAS) summary statistics data released by the International Genomics of Alzheimer's Project Consortium. We used seven single-nucleotide polymorphisms (SNPs) identified to be of genome-wide significance for telomere length as instrumental variables. The MR analysis showed that shorter telomere length was associated with higher odds of Alzheimer's disease.

In Study II, we explored the potential pathways from telomere length to coronary heart disease using network MR method. The same seven SNPs were used as instrumental variables as in Study I. Various GWAS summary statistics data of metabolic biomarkers and coronary heart disease were used. The MR analyses found that shorter telomeres were associated with higher levels of insulin, which also had an effect on coronary heart disease. Overall, this study indicates that insulin lies in the pathway from shorter telomeres to coronary heart disease.

In Study III, we investigated the association of telomere length with trajectories of general cognitive abilities in two Swedish cohorts (SATSA and GENDER) and two US cohorts (MCSA and HRS). Telomere length was measured once, while general cognitive abilities were assessed repeatedly at up to seven occasions. Latent growth curve models were applied to examine the associations. We found that shorter telomere length was associated with lower mean levels of general cognitive ability in the age-adjusted models, but not in models when other covariates were further considered. We did not find evidence to support that telomere length was associated with the decline of general cognitive ability.

In Study IV, we revisited the association of telomere length with all-cause mortality, allowing for time-varying effects in a Swedish twin sample where shared familial confounding could be controlled for. Telomere length was measured using Southern blot method, and data of all-cause mortality was obtained from the Swedish Population Registry. We applied between-within analyses in a shared-frailty generalized survival model framework. We found that shorter telomere length was associated with higher mortality rate when controlling for shared familial confounding. Further, we found significant time-varying effects of telomere length on mortality.

In summary, we presented novel evidence about the role of shorter telomere length in aging related diseases and traits in this thesis. We took advantage of publicly available GWAS summary statistics data as well as individual-level cohort data and used various innovative designs and statistical methods to achieve this. This compiled thesis could contribute substantially to the literature of *the short and long story of telomeres*.

LIST OF SCIENTIFIC PAPERS

I. Yiqiang Zhan, Ci Song, Robert Karlsson, Annika Tillander, Chandra A. Reynolds, Nancy L. Pedersen, Sara Hägg. **Telomere Length Shortening and Alzheimer Disease-A Mendelian Randomization Study**. *JAMA Neurology*. 2015; 72(10):1202-1203.

II. Yiqiang Zhan, Ida K. Karlsson, Robert Karlsson, Annika Tillander, Chandra A. Reynolds, Nancy L. Pedersen, Sara Hägg. **Exploring the Causal Pathway from Telomere Length to Coronary Heart Disease: A Network Mendelian Randomization Study**. *Circulation Research*. 2017; 121:214-219.

III. Yiqiang Zhan, Mark S. Clements, Rosebud O. Roberts, Maria Vassilaki, Brooke R. Druliner, Lisa A. Boardman, Ronald C. Petersen, Chandra A. Reynolds, Nancy L. Pedersen, Sara Hägg. **Association of Telomere Length with General Cognitive Trajectories: a Meta-Analysis of Four Prospective Cohort Studies**. *Manuscript*.

IV. Yiqiang Zhan, Xing-rong Liu, Chandra A. Reynolds, Nancy L. Pedersen, Sara Hägg, Mark S. Clements. **Telomere Length and All-Cause Mortality: A Between-Within Twin Study with Time Dependent Effects Using Generalized Survival Models**. *Manuscript*.

Contents

List of Abbreviations

AD	Alzheimer's Disease
BMI	Body Mass Index
CHD	Coronary Heart Disease
GENDER	Sex Differences in Health and Aging
GRS	Genetic Risk Score
GSM	Generalized Survival Model
GWAS	Genome-wide Association Study
HR	Hazard Ratio
HRS	Health and Retirement Study
LDL	Low Density Lipoprotein
MCSA	Mayo Clinical Study of Aging
MR	Mendelian Randomization
NMR	Network Mendelian Randomization
OCTO-Twin	Origins of Variance in the Old-Old
OR	Odds Ratio
PCR	Polymerase Chain Reaction
SATSA	Swedish Adoption/Twin Study of Aging
SD	Standard Deviation
SNP	Single Neucleotide Polymorphism
TC	Total Cholesterol
TG	Triglyceride
TL	Telomere Length
WC	Waist Circumference

1. Background

Telomeres are repetitive sequence of nucleotides at the end of chromosomes. By protecting the end of chromosomes, they have been recognized as the fundamental aspects of cell biology, especially since the awarding of 2009 Nobel Prize in Physiology and Medicine to Drs. Elizabeth H. Blackburn, Carol W. Greider, and Jack W. Szostak. As an indicator of cellular senescence, the length of telomeres have been postulated as a biomarker of aging in human beings. Although experimental studies have successfully shown that telomere shortening causes cell death or some disorders, results from animal studies do not necessarily translate to or apply to humans. Epidemiological studies play a central role in disentangling the potential influence of telomere shortening or elongation in human health. The causes and consequences of telomere shortening have triggered scientists' curiosity for decades.

The general definition of epidemiology, as a discipline of science, centers on studying the distribution of diseases and associated causal risk factors, diagnosis and prognosis, and treatment evaluation, etc. Modern epidemiological applications, however, span far beyond that. In the aging epidemiology research field, people have shown persistent and increasing interest in the causes of longevity and other chronic diseases and in how to extend life span or to treat aging diseases. Many epidemiological techniques are potentially available for this purpose. For instance, randomized controlled trials are the gold standard to evaluate the effectiveness of a treatment. However, most epidemiological studies use observational data for inference. Empirical evidence is essential, even before initiating a trial.

In this thesis, we take advantage of the modern epidemiological methods and apply them to the publicly available data and our in-house data to examine the potential role of telomere length (TL) in human health. We provide updated epidemiological evidence

of the relationship of TL with coronary heart disease(CHD), general cognitive ability, Alzheimer's disease(AD), and all-cause mortality.

1.1 Telomere Biology

Human telomeres consist of a tract of tandemly repeated short DNA repeats, TTAGGG, and associated protective proteins. Telomeres function as preventing the end of the chromosomal DNA from being fused with other chromosomes that would lead to an unstable genome. Each time a cell divides, telomeres become shorter. This telomere attrition can be replenished by an enzyme, telomerase, that adds telomeric repeat sequences to the ends of chromosomes, hence elongating them to compensate for their loss [1]. Length of telomeres varies across sex, age, and ethnicity. In newborns of different ethnicity, the lengths of leukocyte telomeres were around $9.4 - 11$k base pairs [2, 3]. A previous study observed that TL in nine years old African American boys ranged from $5.0 - 22.5$k base pairs [4]. Women usually have longer telomeres than men [5, 6], and middle aged US adults who self-identified their ethnicity as Black may have longer telomeres than those who self-identified as White [7].

1.2 Measurement of Telomere Length

To date, several methods have been available to measure the length of telomeres. For epidemiological studies with large sample sizes, a fast and reliable technique to assess biomarkers has always been a challenge. Telomeres measurement is not an exception. Besides the measuring techniques, what kind of tissue should be measured is also a concern.

Blood samples are convenient tissues that are commonly collected in most of the cohort studies and can yield a high quality DNA that is available for telomere assays. Other tissues, such as saliva, have also been used for telomeres assessment in epidemiological studies. Previous studies found that TL in blood (leukocytes) has a moderate correlation with that in other tissues [8]. Due to the vast availability of samples, blood leukocyte TL has been generally measured and considered as a biomarker for the overall TL.

Telomere restriction fragment analysis using Southern blot was the first method used to determine the length of telomeres. It measures the absolute length (in base pairs) for specific samples and is a reference for other methods developed afterward [9]. This method requires a relatively large amount of DNA, but has a simple design that does not require specific equipment in a regular laboratory.

Quantitative polymerase chain reaction (PCR) has been developed to assess the length of telomeres in high throughout laboratories [10]. It is fast and requires less DNA than Southern blot; it is also the dominant method in epidemiological studies. Although it is simple and time-saving, it cannot give absolute values of the length of telomeres unless a prediction equation is developed with Southern blot performed on the same sample simultaneously. Thus, it is difficult to compare the results across studies or laboratories

using PCR method. Debate on this continuous. A modified PCR has been developed to measure the absolute length of telomeres [11].

Other techniques to examine telomere length include quantitative-fluorescence in situ hybridization (Q-FISH), fluorescence in situ hybridization and flow cytometry (Flow-FISH), single telomere length analysis (STELA), etc [12, 13]. Q-FISH estimates the cell average length of telomeres and chromosome specific length of telomeres. Flow-FISH measures cell specific length. STELA measures single chromosome end-specific length. More and more methods are being developed now.

1.3 Heritability of Telomere Length

Several studies to date have reported the heritability of TL using twin or family designs. The recent Long Life Family Study with 3037 individuals shows the heritability was 0.54 [14], while an international collaboration study found that heritability of leukocyte TL ranged from 0.62 to 0.8 in different study populations [15]. In Sweden, our group previously showed that TL heritability was 0.56 using data from the Swedish Twin Registry [16]. Different study populations with various age distributions might explain the inconsistent estimates. A Danish study found that the estimated heritability of telomere length declined with increasing age [17].

1.3.1 Genetic Factors Associated with Telomere Length

The first Genome-wide Association Study (GWAS) for TL conducted in 2009 suggested that two variants on chromosome 18q12.2 in the same region $VPS34/PIKC3C$ may influence telomere length, although they did not reach genome-wide significance level [18]. Later, another GWAS with a larger sample size found a variant to be of genome-wide significance in the region of $TERC$, which encodes the telomerase RNA component, and each copy of the minor allele of rs12696304 was associated with 75 base-pair reduction in mean telomere length [19]. A much larger study also confirmed a variant in $TERC$ locus, and additionally discovered a variant in $OBFC1$ as a locus involved in human leukocyte telomere biology [20]. $TERC$ was also confirmed in another study, but $OBFC1$ was not replicated [21]. Later, a collaboration study involving more than 10 000 study participants found two additional loci near $CTC1$ and $ZNF676$ in the general population.

The largest GWAS till now was the ENGAGE Telomere Consortium with 47 000 individuals, and it reported seven SNPs reached genome-wide significance in the close vicinity to : *TERC, TERT, NAF1, OBFC1, RTEL1, ZNF208*, and $ACYP2$ [22]. These findings were replicated in the COGS project, which also revealed a novel variant close to PXK and evidence for TL loci at $ZNF31$ and $BCL2L1$ [23]. As far as I am aware of, another two on-going GWAS investigations (GERA Study and UK biobank) with more individuals will probably find more loci and shed light on telomere biology in the near future.

1.3.2 Environmental Factors associated with Telomere Length

Smoking is a major contributor to adverse health outcomes, and thus has been hypothesized as one of the risk factors of TL attrition as well. Many studies have been conducted to examine this association. In the Nurses' Health Study (NHS), no association between TL and smoking was observed [24], while another study found daily smokers had shorter telomeres than never-smokers [25], which is similar to the findings from the Prevention of Renal and Vascular End-stage Disease Study in the Netherlands [26]. In contrast, the magnitude of the association was much weaker in a German cohort [27], and a longitudinal Danish cohort did not find significant associations [28].

Alcohol consumption has exhibited a U-shaped relationship with regards to some health outcomes. However, such a relationship has not been clearly observed in terms of its association with TL. Although some studies suggest that alcohol consumption was associated with shorter telomeres [29], a cross-sectional study involving 477 individuals observed no significant association between alcohol consumption and TL [25], nor in a Danish cohort [28]. However, studies on alcohol and telomeres are remarkably fewer than smoking; more studies are needed.

Physical activity is always an interesting modifiable factor that deserves investigation in relation to human health. In the Nurses' Health Study (NHS), no association between TL and physical activity was observed [24]. Vigorous physical activity was associated with longer telomeres [25], which was also reported in Berlin Aging Study II that regular physical activity for at least 10 years is necessary to achieve a sustained effect on TL [30]. However, the finding was not significant in a larger sample of Danish population [28].

The Multi-Ethnic Study of Atherosclerosis found no significant association between energy intake and TL [31], similar to those findings in the Nurses' Health Study (NHS) [24]. However, a longitudinal study found more energy intake to be associated with decreased telomeres in Israel [32]. A positive correlation between dietary total antioxidant capacity and telomeres was found, while higher white bread consumption was associated with shorter telomeres in fully-adjusted models in a Spanish cohort [33]. A Chinese study found tea consumption to be associated with longer telomeres. Dietary fiber intake, specifically cereal fiber intake, was positively associated with telomeres [34].

A recent randomized controlled trial reported that supplementation of n-3 fatty acid can elongate leukocyte telomere length in patients with chronic kidney disease in Australia [35]. In a cohort of patients with coronary artery disease, there was an inverse relationship between baseline blood levels of marine omega-3 fatty acids and the rate of telomeres shortening over 5 years [36]. The effect of omega-3 fatty acids on telomere was further confirmed in a randomized controlled trial [37]. Additionally, a significant association between vitamin D and telomeres was observed in US [38]. In the Nurses' Health Study, greater adherence to the Mediterranean diet was associated with longer telomere length [39]. This finding was also reported in a multi-ethnic study, but limited to non-Hispanic whites [40]. A comprehensive review on nutrition and TL can be found elsewhere [41].

The Whitehall II cohort study found a linear association between shorter sleep duration and shorter leukocyte telomeres in men but not in women [42]. But in the Nurses' Health Study, shorter sleep duration was associated with decreased telomeres [43]. Likewise, in a cohort of HIV adults, similar results were obtained that longer sleep at night may either protect telomeres from damage or restore them on a nightly basis [44]. Sleep quality was also reported to associate with shorter telomeres [45], and sleep apnea was found to have similar effect on telomere length [46]. However, another study in an elderly population found insomnia to be associated with shorter leukocyte telomeres in adults aged 70-88 years old, but not in those younger than 70 years [47].

1.4 Telomere Length and Common Diseases

1.4.1 Cardiometabolic Disorders

Compared with risk factors, the associated outcomes of long and short telomeres are more interesting to researchers and the public. Literatures on the association of TL with metabolic disorders including coronary heart disease, diabetes, stroke, etc., are widely available. The relationship between telomeres and CHD is well studied in the literatures and results were mostly consistent across different study populations. A meta-analysis and systematic review summarizing these results found that those with shorter telomeres had higher risks of CHD [48]. Similar results were reported in another meta-analysis, which reported a significant association of shorter telomeres with higher risks of stroke, myocardial infarction, and type 2 diabetes mellitus [49].

Besides non-genetic observational studies, genetic studies also found that people with a higher genetic risk score (GRS) of telomeres constructed using TL-associated variants also had higher risk of CHD [22, 50]. The presumed causal effects of shorter telomeres on CHD seemed to be insensitive to different study designs and populations. Regarding the association between telomeres and stroke risk, more longitudinal studies are needed. A recent Mendelian randomization (MR, method described in the Methods section 3.2.1) study did not find a significant association of shorter telomeres with stroke risk [51].

As for diabetes, a MR study of TL reported no statistically significant association between shorter telomeres and diabetes risk [52]. Sample size and lower power may be a weakness of this study; however, further analysis using DIAGRAM Consortium data did not show a significant association with diabetes, nor with other glycemic traits with the exception for fasting insulin or insulin resistance [51]. This is an interesting finding and deserves further investigation whether the finding could be replicated in future studies with similar sample size and power.

1.4.2 Neurological Disorders

Several studies have reported a significant associations between telomeres and dementia. The Washington Heights-Inwood Community Aging Project found that shortened leukocyte

TL was associated with higher risks of incident dementia [53], similar to that observed in the Medical Research Council-Cognitive Function After Stroke Study [54]. Our study using a MR design and GWAS summary statistics data also obtained similar results [55].

Consistent with these findings, the association between telomeres and cognitive performance were reported in several observational studies [56, 57], and a Mendelian randomization collaboration study of telomeres and general cognitive ability also supported this finding [58]. However, a recent longitudinal study with repeated measurements of both telomeres and cognitive function found that TL changes did not predict changes in cognitive and physical abilities [59].

Studies regarding telomeres and Parkinson's diseases (PD) are much fewer and the findings were inconsistent. A Japanese study did not find significantly shorter telomeres in Parkinson's patients [60]. Similar non-significant results were also reported in the Health Professionals Follow-up Study [61]. A recent meta-analysis summarizing these studies did not report a significant difference for telomeres between 956 PD patients and 1284 controls [62]. Studies of telomeres in relation to other neurological disorders are scarce. A recent study reported shorter telomeres in patients with multiple sclerosis [63] and patients with migraine [64].

1.4.3 Psychiatric Disorders

The relationship between telomeres and psychiatric disorders is much more complicated than other diseases. A large case-control study reported longer telomeres in schizophrenia patients compared with healthy controls [65], while others found the opposite [66, 67]. However, telomeres were reported to be shorter in those with depression as shown in recent meta-analysis [68, 69].

1.4.4 Cancer

Long telomeres have been hypothesized as one of the main causes for infinite cancer cell division for a long time. However, a recent study using genetically predicted telomeres found the story between long and short telomeres and cancers may not be that simple as expected [70]. This study involving five common cancers (prostate cancer, breast cancer, lung cancer, ovarian cancer, and colorectal cancer) using summary data from 51 725 cases and 62 035 controls found that longer telomeres were associated with increases in lung adenocarcinoma risk, but not others [70].

In parallel, another recent comprehensive MR study reported that longer telomeres may be associated with increased risks of colorectal cancer, breast cancer, serous invasive ovarian, lung adenocarcinoma, bladder cancer, chronic lymphocytic leukemia, glioma, and serious low-malignant-potential ovarian cancer; and decreased risks of testicular germ cell cancer and prostate cancer, using a bigger GWAS summary statistics data (21 datasets) [51].

1.4.5 Mortality

The ultimate hypothesis in telomere research that attracts researchers for a life-long time dedication to its study is to elaborate if we can elongate telomeres to extend our life-span. The topic, telomeres and longevity in human being, is very captivating, but also challenging.

Observational studies on telomeres and mortality may offer some clues. The most cited research answering this question really observed some exciting results: people with shorter telomeres had poorer survival, attributable in part to a 3.18-fold higher mortality rate from heart disease and an 8.54-fold higher mortality rate from infectious disease in Utah, USA [71]. However, the Leiden 85-plus Study did not replicate this finding [72], nor did the Lothian Birth Cohort 1921 study [73] or the Zutphen Elderly Study [74]. The MacArthur Health Aging Study further showed that it was TL changes overtime, not baseline telomeres, that could predict mortality [75]. The Bruneck Study in Italy additionally found that shorter leukocyte telomeres were associated with cancer mortality [76, 77]. Findings from the largest study on this topic to date, the Copenhagen City Heart Study and the Copenhagen General Population Study [78], corroborated previous findings. Two more recent studies suggested that these associations could differ by ethnicity [79, 80].

Moving Forward

This chapter has summarized some of the updated results, particularly from epidemiological reports, on telomeres research. These studies have conveyed important and significant information to research community as well as to public society, and played a critical role in evidence synthesis and causal inference on telomeres. They, however, do have some limitations, which could be addressed in future research. In the following chapters, we will examine the role of telomeres in aging-associated diseases and traits and provide additional evidence from new perspectives.

2. Aims

The overall aim of this thesis is to provide evidence for a better understanding of the role of telomere length in the epidemiology of aging. More specifically, the aim of each study is as follows:

I. To examine the causal effect of telomere length on Alzheimer's disease using a Mendelian Randomization approach and GWAS summary statistics.

II. To explore the potential causal pathways from telomere length to coronary heart disease using a network Mendelian Randomization approach and GWAS summary statistics.

III. To investigate the association of telomere length with general cognitive trajectories using data from four cohorts.

IV. To study the association of telomere length with all-cause mortality using generalized survival models.

3. Methods and Results

This chapter was organized as follows. We first described the data and cohorts as well as the variables that were used for Study I-IV in section 3.1, then we introduced the statistical analysis methods in section 3.2. Section 3.3 was presented for results in the order of the four studies.

3.1 Data

3.1.1 GWAS Summary Statistics

The summary statistics were generated by genome-wide association studies, which are observational studies investigating the associations of a set of genome-wide genetic variants with specific traits (eg. telomere length). The publicly available GWAS summary statistics typically focus on the associations of single-nucleotide polymorphisms (SNPs) and traits. In this thesis, the following GWAS summary statistics were used.

Telomere Length GWAS

The telomere length GWAS was conducted by the European Network for Genetics and Genomics Epidemiology (ENGAGE) Consortium [22]. This study consists of 37 684 individuals with replication of selected SNPs in an additional 10 739 individuals. All participants were of European ancestry. Mean leukocyte telomere length was measured by PCR-based technique in all samples. Details of genotyping methods, quality control criteria, and analysis methods were described previously [22].

Alzheimer's Disease GWAS

The Alzheimer's disease (AD) GWAS was performed by the International Genomics of Alzheimer's Project (IGAP) Consortium [81]. The study participants were also of European ancestry. IGAP is a two-stage study, in which 7 million SNPs were meta-analzyed from previously published GWAS datasets consisting of 17 008 AD cases and 37 154 controls in the first stage and 211 632 SNPs were genotyped and tested for the association in an independent sample of 8 572 AD cases and 11 312 controls. Finally, a meta-analysis was performed to combine results from stage 1 and stage 2 [81].

Coronary Heart Disease GWAS

CARDIoGRAMplusC4D (Coronary ARtery DIsease Genome wide Replication and Meta-analysis [CARDIoGRAM] plus The Coronary Artery Disease [C4D] Genetics) consortium represents a collaborative effort to combine data from multiple large scale genetic studies to identify risk loci for coronary artery disease and myocardial infarction. In this thesis, we mainly used CARDIoGRAM GWAS summary statistics as our primary data for analysis because other CARDIoGRAMplusC4D GWAS summary statistics were pooled from studies of both European and South Asian descent, while CARDIoGRAM GWAS is a meta-analysis of 22 GWAS studies of European descent imputed to HapMap 2 involving 22 233 cases and 64 762 controls [82].

Metabolic Biomarkers GWAS

MAGIC (the Meta-Analyses of Glucose and Insulin-related traits Consortium) represents a collaborative effort to combine data from multiple GWAS to identify additional loci that impact on glycemic and metabolic traits [83, 84]. The DIAGRAM (DIAbetes Genetics Replication And Meta-analysis) consortium is a grouping of researchers with shared interests in performing large-scale studies to characterise the genetic basis of type 2 diabetes, and a principal focus on samples of European descent [85, 86]. The Genetic Investigation of ANthropometric Traits (GIANT) consortium is an international collaboration that seeks to identify genetic loci that modulate human body size and shape, including height and measures of obesity [87]. Global Lipids Genetics Consortium performed large-scale meta-analysis of blood lipid fractions (total cholesterol, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, and triglyceride) [88]. Within the ENGAGE Consortium, investigators undertook large-scale meta-analysis of GWAS supplemented by 1000 Genomes imputation for eight quantitative (lipid, glycaemic and obesity-related) traits in individuals of European ancestry [89, 90].

3.1.2 Cohorts

Swedish Twin Registry

The Swedish Twin Registry was established in the 1950s, and becomes one of the largest twin registries in the world. The registry currently contains more than 190 000 twins born between 1886 and 2008. Within the Swedish Twin Registry, several studies were

conducted and are described below.

GENDER

The Sex Differences in Health and Aging (GENDER) is a population-based cohort study from the Swedish Twin Registry and has recruited unlike-sexed twins born between 1906 and 1925 in Sweden [91]. Four hundred and ninety eight individuals of European ancestry participated in the first in-person testing including cognitive tests, health examination, and blood sample collection from 1995. In-person testing follow-up was conducted up to three times on a four-year rolling interval during the years 1995 to 2005 with an average of 5.6 years (standard deviation [SD]: 2.0 years) follow up. In total, four hundred participants had at least one cognitive test and 404 participants had TL assessed. The combined data for TL and general cognitive ability were available for 327 participants.

SATSA

The Swedish Adoption/Twin Study of Aging (SATSA) is a population-based cohort study from the Swedish Twin Registry. It was initiated in 1984 to study twin pairs reared apart or reared together, and 859 individuals of European ancestry participated in at least one wave of in-person testing [92, 93]. In the present study, 632 participants had at least one cognitive assessment during the third, fifth, sixth, eighth, or ninth in-person testing, and 638 participants had telomere length measured. In total, 566 participants had both TL and general cognitive ability assessed. Follow-up was conducted up to 5 times during years 1992 to 2012 with an average of 10.5 years (SD: 5.0 years) follow-up.

OCTO-Twin Study

The Origins of Variance in the Old-Old: Octogenarian Twins (OCTO-Twin) Study is a longitudinal study of twins aged 80 years and over [94] from the Swedish Twin Registry. The first data collection was conducted between 1991 and 1994. Blood samples were collected during the second data collection wave. Twins who were not participants of SATSA were eligible for this study. Blood samples for this study were collected between 1993 and 1996. For the present study, 366 twins born between 1900 and 1928 donated blood samples during the years 1992 to 1996 and had TL measured; this sample has been described previously.

MCSA

The Mayo Clinic Study of Aging (MCSA) is a prospective population-based study using a stratified random sampling design that began in 2004 in Minnesota, US [95]. The study population consisted of participants aged 50 years and above. In this study, 1267 participants had TL assessed and 1225 participants had at least one cognitive measurement. The present analyses included 1205 participants primarily of European ancestry with available data on both TL and at least one cognitive testing. Participants were followed up at 15-monthly intervals for up to seven times from the year 2008 to 2017 with an average of 3.9 years (SD: 1.6 years) follow-up.

HRS

The Health and Retirement Study (HRS) is a nationally representative longitudinal survey of more than 37000 individuals over the age of 50 years in 23000 households in the US. The survey, which has been fielded every two years since 1992, was established to provide a national resource for data on the changing health and economic circumstances associated with aging at both individual and population levels. Details of HRS were described elsewhere [96]. Cognitive ability was assessed up to four times in 20819 participants during years 2008 to 2014 with an average of 5.0 years (SD: 1.5 years) follow-up. A subset of participants (n=5808) had TL assessed. In the present study, we included 3857 participants of European ancestry who had data available on both TL and general cognitive ability.

3.1.3 Measurements

Telomere Length

Telomere length was measured using PCR method in ENGAGE Telomere Consortium, SATSA (Study III), GENDER, MCSA, and HRS, and using Southern blot in SATSA (Study IV) and OCTO-Twin Study. The detailed experimental procedures can be found in a previous article [22] and two manuscripts for Study III and IV.

General Cognitive Ability

In GENDER and SATSA, a general cognitive ability score based on performance on cognitive tests was derived through the extraction of the first principal component analysis of Synonyms, Block design, Thurstone Picture Memory, and Symbol Digit tests, excluding any prevalent dementia cases [97, 98]. The principal component analysis scoring coefficients were applied from the baseline wave to the subsequent waves with subtests z-transformed to their respective baseline means and standard deviations so that intra-individual change could be assessed. For MCSA, a global cognitive z-score was calculated using the z-score-transformed means of the four cognitive domain z-scores for memory, language, executive function, and visuospatial skills domains [99]. For HRS, a total cognitive score was constructed from immediate and delayed word recall, serials 7, backwards counting from 20, and object naming and then z-transformed [100].

All-Cause Mortality

All-cause mortality (Study IV) data with exact dates of death were obtained through linkages of Swedish Twin Registry to the Swedish Population Registry. The mortality data used in this study were updated through November 15, 2017.

Covariates

Educational attainment (Study III) was the self-reported highest education. In GENDER, it was classified as less than elementary school, elementary school, more than elementary school, vocational school, high school, and university or higher. In SATSA, educational

attainment was classified as elementary school, vocational school, high school, and university or higher. Educational attainment was defined as the number of years in school in MCSA, while it was defined as the highest degree of education that was classified as no degree, general education development, high school diploma, two year college degree, four year college degree, Master degree, professional degree (Ph.D., M.D., or J.D.), and degree unknown/some college in HRS. Education Attainment was used as it was reported and was not recoded further in the analyses.

3.2 Statistical Methods

3.2.1 Mendelian Randomization

Mendelian Randomization approach is one of causal inference techniques and adopted from instrumental variable method commonly used in economics and sociology. This approach aims to estimate the association of a variable (eg. exposure variable: TL) with another (eg. outcome variable: CHD) by working with a third variable (eg. genetic variant) as an instrument (named as *instrumentalvariable*) that is independent of the confounders between the exposure and the outcome and avoids reverse causation [101, 102]. If no other biases are present, then the MR estimates have a causal interpretation. MR analysis has become a fashionable method for causal inference in medical research since the last ten years. A quick PubMed search of the term "Mendelian Randomization" or "Mendelian Randomisation" gives the following Figure 3.1.

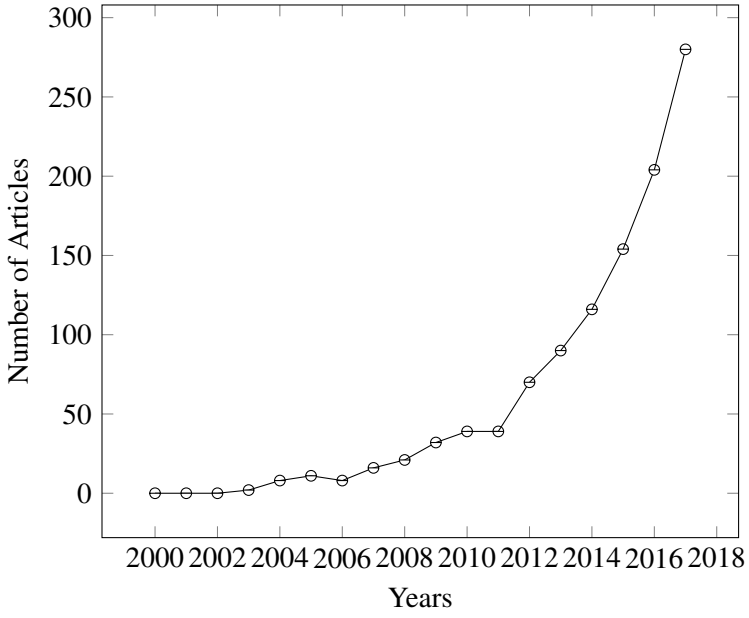

Figure 3.1: Number of Articles in PubMed

The MR approach is usually compared with RCT because it confers some benefits of randomization. The following Figure 3.2 is usually presented to support this argument.

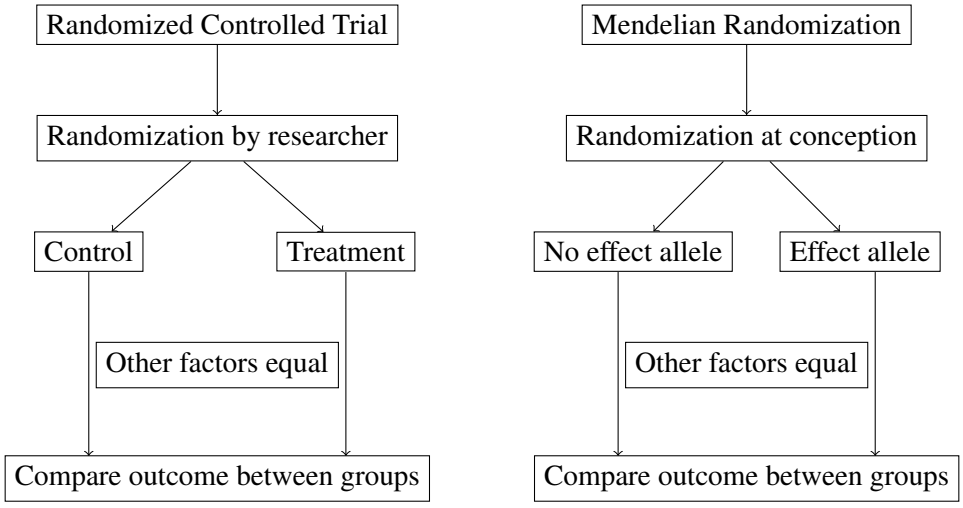

Figure 3.2: Comparision of Randomized Controlled Trial and Mendelian Randomization

The MR framework is illustrated in Figure 3.3. We are interested in testing and estimating the association of TL with some outcome variables. Some measured and unmeasured confounders could bias the results if not well controlled. Suppose we have a genetic variant that only affects TL and there are no other variables other than those drawn in Figure 3.3, then we can conclude that TL is a causal risk factor for the outcome if we find that the genetic variant is also associated with our outcome variable, because the only path that the genetic variant could affect the outcome must be through TL.

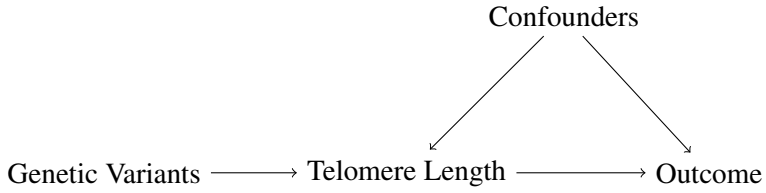

Figure 3.3: Mendelian Randomization Framework

Three assumptions are essential for MR inference:

I. The Genetic variant (eg. SNP) has a causal effect on the exposure variable (eg. TL)
II. The Genetic variant (eg. SNP) affects the outcome (eg. AD) only through its potential effect on the exposure variable (eg. TL)
III. The Genetic variant and the outcome do not have common causes.

These assumptions are sometimes referred to as the instrument *relevance*, *exclusion restriction*, and *unconfoundedness* or *independence* conditions, respectively. The three assumptions are sufficient to test if the exposure (eg. TL) is causally associated with the outcome. The MR estimator will be introduced in Section 3.2.3.

3.2.2 Network Mendelian Randomization

A Network Mendelian Randomization (NMR) extends the MR framework to explore the pathway/mechanism/mediation in a larger system [103]. Figure 3.4 shows an example we used to examine the potential causal pathways from TL to coronary heart disease (Study II). It consists of three different MR tests that are all described below:

I. An MR analysis to determine the causal effect of TL on CHD using TL genetic risk score (GRS) as instrumental variable.
II. MR analyses to examine the causal effects of TL on the metabolic biomarkers (eg. insulin) using TL GRS as instrumental variable.
III. MR analyses to examine the causal effects of the possible mediators (eg. insulin) on CHD using metabolic biomarker GRS as instrumental variable.

If causal associations are observed in all three steps, then a conclusion can be drawn that the specific metabolic risk factor may lie in the causal pathway from TL to CHD.

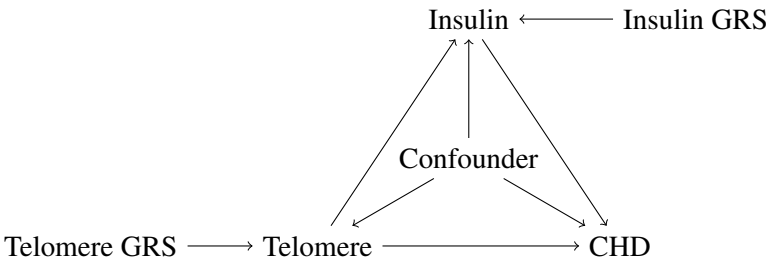

Figure 3.4: Network Mendelian Randomization

3.2.3 MR Using GWAS Summary Statistics

MR is an appealing approach to draw causal inference and does not need to measure confounders (SNPs, exposure, outcome are sufficient in theory). However, it does require a large sample size. Applying MR in practice thus becomes infeasible for empirical epidemiologists because our sample sizes are usually not big enough for MR analysis. One solution is to take advantage of the GWAS summary statistics data, which have offered a unique opportunity to perform MR or NMR [104].

For a single genetic variant as the instrumental variable, the common MR estimator, Wald ratio estimator can be used. It is essentially the ratio of association of the genetic variant with the outcome ($G-Y$, e.g. SNP-CHD) and the association of the genetic variant with the exposure ($G-X$, e.g. SNP-TL):

$$\beta_{iv} = \frac{\beta_{GY}}{\beta_{GX}} \tag{3.1}$$

where β_{iv} is the MR estimate, β_{GY} denotes the association of genetic variant with the outcome, and β_{GX} represents the association of genetic variant with the exposure. The value of β_{GX} can be found in published articles and that of β_{GY} can be obtained from

GWAS summary statistics data or from the authors of that GWAS article where the outcome data are from.

When multiple genetic variants are available as the instrumental variables, special methods are needed to combine these multiple genetic variants. The first and most commonly used method is the inverse variance weighted approach [105], where the final MR estimator is viewed as the weighted average of estimates from each variant:

$$\beta_{iv} = \frac{\sum_{i=1}^{n} \beta_{GX}\beta_{GY}\sigma_{GY}^{-2}}{\sum_{i=1}^{n} \beta_{GX}^{2}\sigma_{GY}^{-2}} \tag{3.2}$$

where β_{iv} is the MR estimate, β_{GY} and σ_{GY} denote the association of the genetic variant with the outcome and standard error, and β_{GX} represents the association of the genetic variant with the exposure.

Other estimators, such as MR-Egger regression and weighted median method are also available for MR inference, and can be used as the alternative methods for sensitivity analysis [106, 107]. We did not use them here in this thesis.

3.2.4 Latent Growth Curve Models

Latent growth curve models were developed to model repeatedly measured variables. These models can be fitted via both structural equation modelling approaches and multilevel modelling methods [108]. They are also referred to as mixed models or random coefficients models, which are common in medical literature. In this thesis (Study III), we used the random effects model and attained age was centered. A couple of models were fitted and compared in Study III.

The basic random intercept and slope model to assess the association of TL with the mean levels of general cognitive ability is specified as:

$$y_{ij} = \beta_0 + u_{0i} + (\beta_1 + u_{1i})age_{ij} + \beta_2 x_i + \beta_3 sex_i + \varepsilon_{ij} \tag{3.3}$$

where i represents the individual and j is the measurement occasion, x and y denote telomere length and general cognitive ability, u_{0i} and u_{1i} are random intercepts and slopes, respectively, and they follow $\left[\begin{smallmatrix} u_{0i} \\ u_{1i} \end{smallmatrix}\right] \sim MVN\left(\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{smallmatrix}\right]\right)$, the random error term $\varepsilon_{ij} \sim N(0, \sigma^2)$.

To examine the association of TL with the decline of general cognitive ability, the following model is specified as:

$$y_{ij} = \beta_0 + u_{0i} + (\beta_1 + u_{1i})age_{ij} + \beta_2 x_i + \beta_3 sex_i + \beta_4 x_i age_{ij} + \varepsilon_{ij} \tag{3.4}$$

where i represents the individual and j is the measurement occasion, x and y denote telomere length and general cognitive ability, u_{0i} and u_{1i} are random intercepts and slopes, respectively, and they follow $\left[\begin{smallmatrix} u_{0i} \\ u_{1i} \end{smallmatrix}\right] \sim MVN\left(\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{smallmatrix}\right]\right)$, the random error term $\varepsilon_{ij} \sim N(0, \sigma^2)$.

3.2.5 Generalized Survival Models

Generalized survival models (GSM) were developed for survival data analysis [109, 110]. It has the advantage of estimating the baseline hazard and effect measurements in a more flexible and elegant fashion. It can be easily extended to accommodate the between-within twin decomposition, which allow us to control for unmeasured shared confounders for twins within the same twin-pair [111, 112]. In this thesis (Study IV), we used the between-within twin shared frailty model in the GSM framework to study the association of TL with all-cause mortality.

For the thesis, the between-within shared frailty model in the GSM framework was specified with a log-log link function. For a time-constant effect, the model was specified as:

$$\log\Big(-\log\big(S(t_{ij}|x_{ij},u_{ij})\big)\Big) = \log(u_{ij}) + \beta_1 sex_i + \beta_B \bar{x}_i + \beta_W x_{ij} + s_0(t_{ij};\gamma) \tag{3.5}$$

where i represents the twin-pair, j is the individual twin, the frailty term u_i is assumed to follow a gamma distribution, x_{ij} denotes TL for each individual, β_B is the between twin-pair effect, $\bar{x}_i$ is the mean TL in each twin-pair, β_W is the within twin-pair effect, and $s_0(t_{ij};\gamma)$ is a smooth function representing the transformed baseline survival function. This model with a log-log transformation of survival is interpreted as a proportional hazards model. For time-dependent hazard ratios, the model was specified as:

$$\log\Big(-\log\big(S(t_{ij}|x_{ij},u_{ij})\big)\Big) = \log(u_{ij}) + \beta_1 sex_i + \beta_B \bar{x}_i + x_{ij} s_1(t_{ij};\beta_W) + s_0(t_{ij};\gamma) \tag{3.6}$$

where $s_1(t_{ij};\beta_W)$ is a smooth function representing the time-dependent within twin-pair effect and where β_W is now a vector of parameters. The smooth parameters were modeled as natural splines with 2 degrees of freedom for baseline survival and 1 degree of freedom for the time-dependent hazard ratio; the degrees of freedom were selected using AIC and the associations were robust to the choice of degrees of freedom. The models were fitted using maximum marginal likelihood estimation, with suitable adjustment for left-truncation under a shared frailty model [113]. We compared models with and without the time-dependent effects using a likelihood ratio test. We also assessed whether the effect of TL was linear by including the natural spline term for TL in the time-constant model and compared the models with a likelihood ratio test. The within effect, β_W is what we are interested in.

3.2.6 Statistical Packages

All analyses were conducted using SAS 9.4 (SAS Institute, Cary NC) and R 3.4 [114]. Meta-analysis was performed using the *rmeta* package [115]. Heterogeneity test was conducted in the *gtx* package [116]. Figures were plotted using the *ggplot2* package [117]. The GSMs were implemented in the *rstpm2* package [118].

3.3 Results

This section is organized as follows. Study I and Study II used similar methods and are presented first. The results for Study I is described in Section 3.3.1, and those for Study II are delineated in sections 3.3.2, 3.3.3, and 3.3.4. Section 3.3.5 is for Study III, and the last section 3.3.6 is presented for Study IV.

In Study I and Study II, we selected seven SNPs (rs10936599, rs2736100, rs7675998, rs9420907, rs8105767, rs755107, and rs11125529) which are of genome-wide significance for TL as instrumental variables. These SNPs are located or near to *TERC, TERT, NAF1, OBFC1, ZNF208, RTEL1, ACYP2*. They are identified from a previous GWAS of TL in the ENGAGE Telomere Consortium [22]. Theses SNPs and their associated effect sizes on TL are listed in Table 3.1.

SNPs	**Gene**	**Risk Allele**	$\beta(se)$
rs10936599	*TERC*	T	$-0.097(0.008)$
rs2736100	*TERT*	A	$-0.078(0.009)$
rs7675998	*NAF1*	A	$-0.074(0.009)$
rs9420907	*OBFC1*	A	$-0.069(0.010)$
rs8105767	*ZNF208*	A	$-0.048(0.008)$
rs755017	*RTEL1*	A	$-0.062(0.011)$
rs11125529	*ACYP2*	C	$-0.056(0.010)$

Table 3.1: Selected SNPs for Telomere Length and Effect Sizes

3.3.1 Telomere Length and Alzheimer's Disease

In study I, we examined whether TL was associated with Alzheimer's disease by applying MR approach to the GWAS summary statistics for Alzheimer's disease from the IGAP Consortium [55, 81]. We extracted the effect sizes and associated standard errors for the seven SNPs as listed in Table 3.1 from the IGAP Consortium, and then harmonized them to align the effect alleles to be the same for both TL and AD. We applied inverse-variance weighting approach for the MR analysis.

By combining the seven SNPs shown in Table 3.1 as the instrumental variables, the MR analysis showed that one standard deviation (SD) *decrease* of TL was associated with higher odds of AD (odds ratio: 1.36, 95% confidence interval: 1.12-1.67). We also performed the heterogeneity test, which did not show a significant heterogeneous effects among the seven SNPs ($P = 0.29$). These SNPs and their associated effects on AD and TL were visually plotted in Figure 3.5.

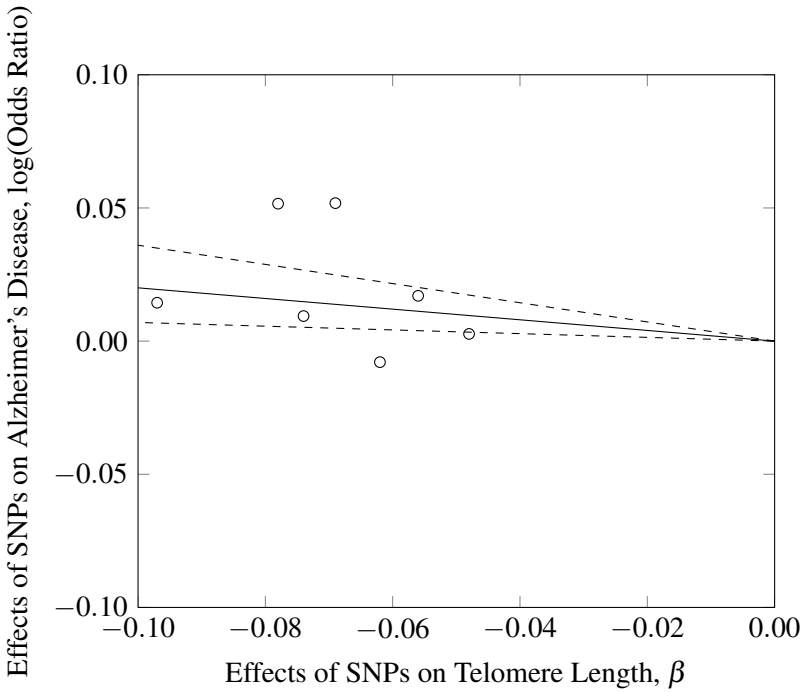

Figure 3.5: Plot of the Effect Sizes of SNPs on Telomere Length and Alzheimer's Disease. The slope of solid line denotes the effect of one SD *increase* of TL on AD: log(odds ratio) and those of dashed lines represent 95% confidence intervals.

3.3.2 Telomere Length and Coronary Heart Disease

Study II aimed to explore the potential causal pathway from shorter TL to CHD using the network MR approach. In the first step, we replicated previous findings which reported that shorter TL was a causal risk factor for CHD from the MR perspective. The MR analysis using the same seven SNPs shown above as the instrumental variables showed one SD *decrease* of TL was associated with 1.26 (95% CI:1.03, 1.54) higher odds of CHD in the CARDIoGRAM Consortium. The effects of these SNPs on CHD were described in Table 3.2 and Figure 3.6. A previous analysis oriented from the genetic risk score analysis was performed in the ENGAGE Telomere Consortium. Our result is similar to that reported.

SNPs	**Gene**	**Risk Allele**	$\log(OR), se$
rs10936599	*TERC*	T	0.026(0.016)
rs2736100	*TERT*	A	0.021(0.017)
rs7675998	*NAF1*	A	−0.025(0.017)
rs9420907	*OBFC1*	A	0.058(0.021)
rs8105767	*ZNF208*	A	0.019(0.016)
rs755017	*RTEL1*	A	0.004(0.023)
rs11125529	*ACYP2*	C	0.020(0.019)

Table 3.2: Association of SNPs for Telomere Length with Coronary Heart Disease.

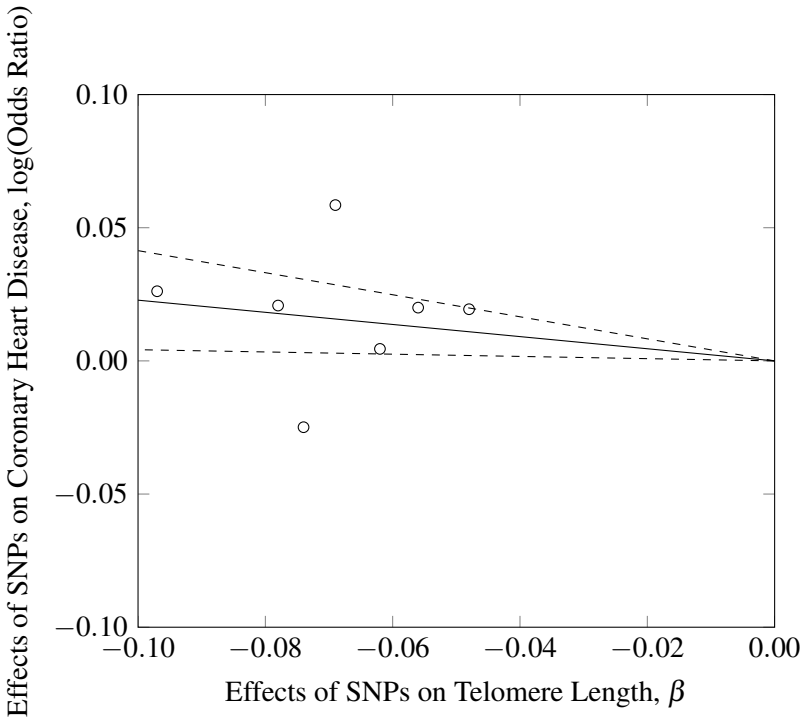

Figure 3.6: Plot of the Effect Sizes of SNPs on Telomere Length and Coronary Heart Disease. The slope of solid line denotes the effects of one SD *increase* of TL on CHD: log(odds ratio) and those of dashed lines represent 95% confidence intervals.

3.3.3 Telomere Length and Metabolic Biomarkers

The MR estimates of TL with metabolic biomarkers are presented in Table 3.3. These biomarkers include fasting insulin, triglycerides, total cholesterol, low-density lipoprotein, fasting glucose, diabetes mellitus, HbA1c, body mass index, waist circumference, and waist to hip ratio. Because several large GWAS summary statistics data from different consortia were available for the same biomarker, we used the earlier and largest GWAS summary statistics as the discovery data and the most recent data as the replication data. Our MR analyses showed that TL was associated with fasting insulin and lipid fractions in the discovery phase. However, only the result for fasting insulin was replicated in the replication phase. One SD *increase* of TL was associated with -0.07 unit decrease of fasting insulin.

3.3.4 Insulin and Coronary Heart Disease

Based on the previous MR examination of TL with metabolic biomarkers, fasting insulin was suggestive of being a potential mediator from shorter TL to increased risk of CHD. Thus, we further evaluated whether insulin was associated with CHD using MR analysis. We used 12 SNPs that were reported by the MAGIC Consortium to be associated with fasting insulin as the instrumental variables (Figure 3.4). We found one unit *increase* of log-transformed fasting insulin levels was associated with higher odds of CHD (OR:

Metabolic Biomarkers	β	*se*
Discovery		
Fasting Insulin	−0.0679	0.027
TG	0.0796	0.027
LDL	0.0746	0.030
TC	0.0702	0.029
Fasting Glucose	0.0236	0.026
Diabetes	−0.1205	0.122
HbA1c	0.0018	0.023
BMI	−0.0005	0.023
WC	0.026	0.027
WHR	0.050	0.026
Replication		
Fasting Insulin	−0.0806	0.029
TG	0.0386	0.039
LDL	0.0736	0.039
TC	0.0453	0.039
WHR	0.014	0.041

Table 3.3: MR Estimates for the Association of Telomere Length with Metabolic Risk Factors for Coronary Heart Disease. TG: triglyceride, LDL: low density lipoprotein, TC: total cholesterol, BMI: body mass index, WC: waist circumference, WHR: waist to hip ratio.

2.61, 95% CI:1.55-4.48). Because rs1421085 is located in the FTO locus and potentially pleiotropic, it was excluded further. Analysis of the remaining SNPs yielded an effect size of 2.41 (95% CI: 1.31-4.26). Additionally excluding rs2972143 in $IRS-1$ revealed an OR of 1.86 (95% CI: 1.01-3.41).

Hence, based on these analyses above, we found that TL was a causal risk factor for CHD in the first step, and was a causal risk factor for insulin which was further a causal risk factor for CHD. Taken together, these analyses suggest fasting insulin might act as a mediator in the causal pathway from shorter TL to CHD.

3.3.5 Telomere Length and General Cognitive Ability

In Study III, we applied the latent growth curve models to data from 5955 participants of two Swedish cohorts (GENDER and SATSA) and two US cohorts (MCSA and HRS), and found that one SD *increase* of TL was associated with 0.021 unit increase of standardized mean general cognitive ability (95% CI: 0.001,0.042) after adjusting for attained age. The point estimate remained similar (0.019) but with wider confidence intervals (95%: −0.002, 0.039) after further adjusting for sex. The association was attenuated with additional adjustment for educational attainment, depression, and coronary heart disease. We did

SNPs	Gene	Risk Allele	$\log(OR), se$
rs2820436	*LYPLAL1*	C	0.023(0.015)
rs1530559	*YSK4*	A	0.012(0.020)
rs10195252	*GRB14*	T	0.014(0.014)
rs2972143	*IRS1*	A	0.052(0.014)
rs9884482	*TET2*	C	0.018(0.015)
rs4865796	*ARL15*	A	0.034(0.014)
rs2745353	*RSPO3*	T	0.003(0.014)
rs1167800	*HIP1*	A	0.027(0.016)
rs983309	*PPP1R3B*	T	−0.031(0.024)
rs7903146	*TCF7L2*	C	−0.019(0.015)
rs1421085	*FTO*	C	0.031(0.015)
rs731839	*PEPD*	G	0.021(0.016)

Table 3.4: Association of 12 SNPs for Insulin with Coronary Heart Disease.

not find strong evidence for the association of TL with decline (i.e. the slope) of general cognitive ability ($\beta = 0.002$, 95% CI: -0.002, 0.004).

3.3.6 Telomere Length and All-Cause Mortality

In Study IV, we applied the between-within shared frailty model to the twin data (SATSA and OCTO-Twin). Of the 366 twins, 115(31.4%) were men and 251 (68.6%) were women. The mean ages of the participants at blood drawn were 76.5 years for men and 80.3 for women. Follow-up duration spanned from ten days to 25.7 years (mean: 10.2 years). During the follow-up period, 341(93.2%) participants died.

We found that shorter TL was associated with higher mortality rate in general (within twin-pair effect). Assuming a time-constant within effect, the estimated hazard ratio for a 0.5 kbp *reduction* in TL was 1.18(95% CI: 1.00, 1.38). We did not find evidence for a non-linear effects of TL on all-cause mortality (likelihood ratio test, $P = 0.32$). There was, however, strong evidence that this association was not time-independent (likelihood ratio test, $P = 0.001$). The time-dependent hazard ratios approached 1 with increasing age, whereas the point-wise hazard ratios were not statistically significant at around 90 years old (Figure 3.7). We also performed additional analyses to examine if results differed by sex, zygosity, or varying degrees of freedom of the smooth function. These analyses, however, yielded similar results.

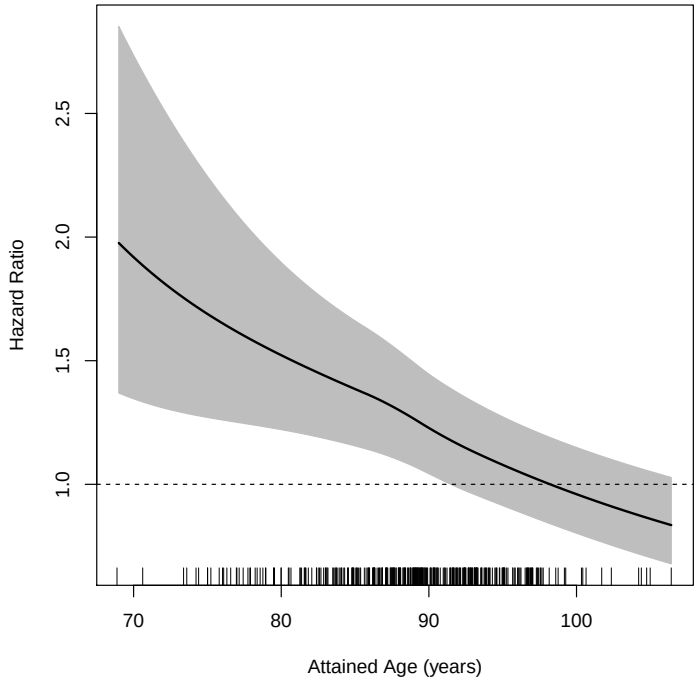

Figure 3.7: Association of Telomere Length with All-Cause Mortality. The solid line represents the hazard ratio for 0.5 kbp (kilo base pairs) *decrease*, and the grey shade denotes the point-wise 95% confidence intervals. The rug on the x-axis represent the event times

4. Discussion and Conclusion

The work in this thesis has examined the associations of leukocyte telomere length with several outcomes including coronary heart disease, metabolic biomarkers, Alzheimer's disease, general cognitive ability, and all-cause mortality, which are among the most commonly studied traits in epidemiological research on aging. We found that telomere length was associated with coronary heart disease, Alzheimer's disease, and all-cause mortality, but not with general cognitive ability. Furthermore, we found that insulin might mediate the association of telomere length with coronary heart disease. These findings added new evidence in the literature of *the short and long story of telomeres*.

4.1 General Discussion

4.1.1 Implications

The four studies in this thesis highlight the importance of telomeres in the etiology of aging-related diseases and traits, specially CHD, insulin, AD, and mortality. These findings largely corroborate previous results of the same topics. However, our studies examined the role of telomeres in aging from novel perspectives. In Study I and II, we applied MR method, the results from which are independent of measured and unmeasured confounding and reverse causation. In Study III, we used longitudinal data, while in Study IV we exploited twin data. If the modelling assumptions hold and no other biases are present, then shorter telomeres are causes of CHD, insulin, and AD. However, the potential roles of telomeres in clinical setting or drug development are beyond the scope of this thesis. The following discussions might serve as an introduction to the unanswered questions.

4.1.2 Telomere Length Measurement Error

As repeated sequence of nucleotides, telomeres are difficult to measure in large-scale epidemiological studies. Measurement error challenges the validity of epidemiological research and makes comparison of results across different studies troublesome.

In Study I and II, we only used the TL−associated SNPs as proxies for TL, and did not have the physically measured TL. Using TL−SNPs as instrumental variables can minimize the impact of its measurement error (non-differential). This is one of the advantages in MR analysis [119]. In the ENGAGE Telomere Consortium, the measurement error in TL does not affect the final estimates (effect sizes: βs) but affects the associated standard errors. Fortunately, the sample size in ENGAGE Telomere Consortium is large, and thus these standard errors of the selected top SNPs are small. We thus could assume that the effect sizes of these SNPs are valid. We then used these SNPs (with βs) in the MR analyses followed.

In Study III and IV, we used the measured TL as the exposure variable and examined its association with general cognitive ability and all-cause mortality. In these scenarios, measurement error of TL does affect the estimates of these associations. The observed association is generally underestimated and depends on the proportion of the true variance explained by the observed variance of measured TL.

4.1.3 Mendelian Randomization Assumptions

The validity of MR inference relies on the three core assumptions. Assumption I, *instrument relevance*, can be empirically tested by examining the associations of SNPs with TL. Assumption II, *unconfoundedness* or *independence*, is assumed to hold by the nature of Mendelian's law (population stratification may still be a confounder which has been adjusted for in the GWAS analysis, thus this could be called as *conditional independence*). The main concern in MR literature centers on Assumption III, *exclusion restriction*, which says that the effects of SNPs on the outcome are totally through the exposure (TL) or these SNPs have no pleiotropic effects on other traits. Assumption III cannot be tested empirically.

Besides the three core MR assumptions, additional assumption is necessary for the analysis using GWAS summary statistics. The estimates of SNPs' effects on the exposure (TL) and on the outcomes are from different GWAS consortia, these different study populations are assumed to be representative of the same underlying general population (European ancestry in this thesis). Otherwise, the MR results are biologically impossible and may not make sense.

In the MR literature, a lot of efforts have been made to improve the analysis, particularly when Assumption III could be violated. The MR-Egger regression was developed to test the pleiotropy and to correct for it by adding an intercept term in the weighted linear regression models [106]. This approach is usually underpowered and taken as a sensitivity analysis in practice. Other approaches include weighted median method [107], robust

adjusted profile score method [120], etc.

In Study I and II, we did not use these alternative methods. One reason is that these alternative methods are underpowered. A wide confidence interval does not help much in achieving our primary aims. Another reason is the number of SNPs we used in Study I and II is small for these methods. Our view is that a sufficient number of SNPs were necessary for these methods. Unfortunately, we did not find an article discussing the number of SNPs in MR analysis by the time of conducting Study I and II. Our results relied on the inverse-variance weighting method.

The selection of SNPs for MR analysis was based on genome-wide significance levels. This selection method is common in MR practice, particularly when the mechanisms of these SNPs with TL are generally unknown. Selected SNPs with unknown function based on this selection method are prone to pleiotropy. However, some of these selected SNPs in this thesis are located in *TERC*, *TERT*, *RTEL1*, *NAF1*, and *OBFC1* genes, which are involved in telomerase or telomere biology [22]. The potential biological function with these SNPs lays foundation to justify the validity of these SNPs.

One of the advantages of MR analysis is that the investigators do not need to measure the confounders between the exposure and the outcome. This advantage is such appealing that sometimes the investigators do not even think about what the potential confounders could be. Which confounder or confounders do we want to control for but cannot do so and then turn to MR approach to achieve this due to various reasons? In our study, we would like to control for the inflammatory biomarkers and familial confounding that cannot be sufficiently adjusted for in empirical analysis using cohort data.

MR analysis using summary statistics is a brilliant initiative to maximize the utility of GWAS results. However, the secondary analysis of these GWAS summary statistics comes with cost. Due to the limitations, the MR analysis should not be over-interpreted in practice. It can be used primarily for testing purpose in which the associations between SNP and the outcome could be tested in the first place. If a significant finding is observed, then it is possible that the genetically determined exposure or the exposure *per se* is associated with the outcome. If a null finding is obtained, then it is less possible that the exposure is associated with the outcome. In general, a null finding is more plausible than a significant finding if sample size and statistical power are not concerns [121].

4.1.4 General Cognitive Ability

A previous MR study using the same seven SNPs as instrumental variables examined TL and its association with cognitive performance and found that shorter TL was associated with lower levels of general cognitive performance [58]. In the same study, however, the observational analysis did not find a significant association, which was similar to our results. The reasons may lie in the fact that in both these studies, telomeres were measured using PCR that was prone to measurement error as discussed previously. Another reason might be that the true magnitude of the association of TL with general cognitive ability

is indeed small. Taken together, estimating a small effect size with an exposure variable measured with error requires a large sample size.

4.1.5 All-Cause Mortality

The relationship between TL with all-cause mortality has been examined extensively in the literature. Most of the studies found shorter telomeres to be associated with higher mortality rate, while a few others did not observe this association. The differences in study population, telomere measurement methods, and variables controlled in the design or analysis could account for the heterogeneity of results.

Although a number of confounders have been adjusted for in previous studies, residual confounding such as family background or genetics could still bias these results. An elegant way to control for shared familial confounding is to exploit the twin data, in which the twin-pairs share familial characteristics. By controlling for the familial confounding and using the between-within twin design and generalized survival models, we still observed time-dependent association of TL with mortality rate. Taken together, these findings indicate that the association of TL with all-cause mortality were independent of these potential confounders.

4.2 Strengths and Limitations

4.2.1 Strengths

Studies compiled in this thesis provided valuable knowledge in the telomere research literature, particularly for the aging epidemiology community. These studies are based on novel statistical methods, the GWAS summary statistics data, and individual-level data of good quality.

The MR approach allows us to examine the causal effects of TL on a number of outcomes, while the GWAS summary statistics data are created from large collaboration cohorts and offer a unique opportunity to apply the MR method. Besides that, the GWAS summary statistics also provide us with an alternative resource to elaborate the pathways, especially from a biomarker to a disease. In our study, the biomarker is telomere length. The biomarker could also be other variables, such as educational attainment, which is of greater interest for social scientists.

In Study III, we obtained data from repeated measurements of general cognitive ability up to seven occasions. One advantage of longitudinal data is that they could be used to examine the trajectory of a continuous variable and associated factors with its decline. In this study, we did not find a strong evidence to support TL to be associated with cognitive decline, while the magnitude of the association of TL with mean general cognitive ability was small.

In Study IV, telomere length was measured using Southern blot method, which is regarded as the gold standard for telomere length assessment. Then Southern blot method

is costly and requires large amount of DNA. There are fewer epidemiological studies using Southern blot than that using the PCR method. Telomere length reported in absolute length (number of base pairs of nucleotides) has facilitated the interpretation of final estimates. Besides that, the relatively more accurate measurements could minimize bias.

The application of between-within twin study using the generalized survival models is novel. Confounding control is one of the central topics in epidemiology and other quantitative disciplines. Many other techniques are also available for confounding control. The successful application of the between-within model in Study IV allows us to examine the relationship of two variables from a new perspective. Taking into account of previous findings using different analytic methods on this topic, we are more confident about the relationship between TL and all-cause mortality when a variety of methods pointed to the same conclusion.

4.2.2 Limitations

Study I and II used MR approach, the data are from the GWAS summary statistics. Besides the MR assumptions, some other limitations should be acknowledged. Because the individual-level data are unavailable, more analyses cannot be performed. We cannot examine the gene-environment interaction, the non-linear relationship, nor stratified analyses (for example, by sex, which is a common interest in medical research).

In addition to the measurement error for TL in Study III, general cognitive ability was defined differently among the four cohorts. Measuring general cognitive ability in a consistent and detailed way could definitely empower the inference. However, to the best of our knowledge, even both TL and cognitive function are to be measured without error, their association are not expected to be large. In this study, only a limited covariates were considered as confounders and adjusted for. However, due to the null findings and small magnitude of the associations, additional adjustment for other covariates could hardly change the results or conclusions.

Study populations in this thesis are the elderly, particularly in Study IV where the entry age of these participants were around 80 years old. Although the primary aim of this thesis is examining telomere length in aging from the epidemiological perspective, it is very interesting to revisit these analyses in a younger population from the life course perspective.

Another common limitation to most epidemiological studies for telomeres is that telomeres are measured as the average across all chromosomes within a cell and a mixture of leukocytes including neutrophils, lymphocytes, and monocytes, etc. Thus, it is not quite clear which telomeres are affecting the outcome. This concern can be alleviated if we use TL-associated SNPs as proxies, assume these different telomeres are highly correlated, and are interested in telomeres in general. However, more accurate measurement of TL in different cells and chromosomes may provide further insights into the biological mechanisms of telomeres in the aging process.

4.3 Future Perspectives

The consequences of telomere shortening in human beings still need to be investigated. Increasing knowledge about the genetics of telomeres has resulted in valuable insights into telomere biology and its related diseases. However, the known telomere genes only account for a small proportion of total variance in telomere length. In addition to that, whether leukocyte telomere length reflects global or tissue specific telomere length still deserves to be examined. Although exciting results were observed for the association of leukocyte telomeres with many diseases, questions remain to be answered regarding if it is leukocyte telomeres *per se* or together with tissue specific telomeres that contribute to the disease development. Besides these etiological examinations, additional studies looking at the diagnostic or prognostic performance of telomeres (regardless of leukocyte or tissue-specific) may also be interesting from the clinical perspective in the near future.

The causes and trajectories of telomere shortening are not studied in this thesis. Due to current limitations in telomere measurement, alternative methods (such as Mendelian randomization using GWAS summary statistics) could provide valuable insights into the causation of telomere shortening as well as examining if telomeres are involved in the pathway from the variable of interest (eg. smoking) and the outcome of interest (eg. coronary heart disease). Longitudinal studies with repeated measurements of telomeres could also shed light on telomere dynamics. These investigations are definitely essential for the completion of *the short and long story of telomeres*.

4.4 A Note on Causal Inference

Causal inference is one of the central tasks that modern epidemiology aims to complete. We, epidemiologist, are eager to hunt and hopefully find some factor(s) that can cause some diseases. We also know, even not trained in school, that a correlation between two variables is true in that it correctly reflects these two are associated, but does not necessarily mean one causes the other. Anaxagoras, an pre-Socratic Greek philosopher, said *All things were together, infinite both in number and in smallness*. However, a correlation or association between two variables is usually not what we are interested in, how they are associated/connected is. Are they associated because of something else or because one causes another? If something else, a third factor or a set of factors, might causes both X and Y, the influence of the third factor(s) should be eliminated. This is confounding control that has been being practiced all the time in epidemiology.

Many techniques are available to control confounding, such as randomization in the study design stage or adjustment in the analysis stage. MR and within-between analyses are also two of them to achieve this. Usually we reply on regression models with assumptions that these variables are sufficient to control for confounding and the models are correctly specified. Indeed, if this is the true data-generating mechanism and a significant association is observed, a causal conclusion can be drawn. However, we tend not to do so or even not

to use the word *causal* in our papers. Why? We know the assumptions are probably not satisfied: residual confounding is a concern in most scenarios; the reviewers, editors, and other researchers also know this and criticize more. Then a question naturally comes: how can we conclude *causal* in practice?

Besides randomized controlled trials, causal inference from observational studies, in my view, first requires replication. Replicating previous findings in the same underlying population using the same design and analytic methods is necessary. Then replicate them in different populations, and/or using different designs and/or different analytic approaches. For example, the research questions in this thesis, are shorter telomeres (causally) associated with aging-associated diseases and traits (AD, CHD, general cognitive ability, and mortality), were addressed previously in different settings. We have now examined them using novel approaches and from different perspectives. Taking all these evidence together could definitely result in more robust insights in causal inference.

4.5 Concluding Remarks

The telomere story could be longer, but the thesis is short. In this thesis, several novel analytical methods that have covered Mendelian randomization, latent growth curve models, between-within twin study, and generalized survival models were applied to evaluate the role of telomeres in aging. The four pieces of studies found shorter TL was generally associated with worse health outcomes.

New evidence supporting shorter telomere length with higher odds of Alzheimer's disease were presented, novel mechanisms for shorter TL and coronary heart disease were proposed, and the fundamental interest of TL shortening and all-cause mortality was revisited from new perspectives. These findings deserved to be replicated in different and new settings, such as non-European populations. Although we did not find strong evidence underpinning the relationship of telomeres and general cognitive ability, further studies with better measurements are warranted. In particular, telomere length measurement techniques deserve to be improved and tailored for aging epidemiological research.

ACKNOWLEDGEMENTS

The life-changing PhD study has truly been an unforgettable experience. It would not have been possible without the support and advice that I have received through the past four years. I would like to first thank all study participants, data collection staff, and investigators for their efforts to create such precious data and cohorts. I would also like to thank special people who have shared their knowledge and offered generous help to make this thesis different:

Sara Hägg, my main supervisor, for taking me as your first PhD student, supporting me in various ways, and providing me with a fantastic platform to train me as an independent researcher. In addition, I appreciate the very interesting topics you proposed for this thesis and the insightful advice on the written works. You are such a nice person and I am very lucky and grateful to be your student.

Nancy L. Pedersen, my co-supervisor, for your constant support, guidance, and motivation. I really admire your scientific attitude to do good science rather than to focus on impact factor. I appreciate your help in editing our manuscripts and this thesis during your busy schedule. **Chandra A. Reynolds**, my co-supervisor, for investing your sleep time in the project discussion very early in the morning, and for inviting me to visit your lab a few times, although I have not made it even once.

Scott Montgomery, my mentor, for the occasional talks and invaluable advice on work and life.

Mark S. Clements, **Robert Karlsson**, and **Annika Tillander**, for sharing your expertise in biostatistics and programming. Especially for Mark, it is you who proposed the fourth study that has made this thesis more captivating. I can still remember the times you asking me to sit at your desk and revising my manuscript with explanations on hard paper and humour in the air.

Ci Song, **Ida K. Karlsson**, **Xing-rong Liu**, for your helpful comments and suggestions to improve our manuscripts. It is you guys who have pushed our manuscripts to move forward and to be in their current forms. My collaborators in Mayo Clinic: **Rosebud O. Roberts**, **Maria Vassilaki**, **Brooke R. Druliner**, **Lisa A. Boardman**, and **Ronald C. Petersen**, for generously sharing your data and knowledge for our project. The **Health and Retirement Study**, for making the data open access.

Sara and Nancy's group members and other MEB colleagues, for making MEB such an enjoyable workplace.

My parents and **Qing**, for your endless love, trust, and support throughout my life.

This work was funded by a Karolinska Institutet delfinansiering (KID) grant for doctoral student, the Loo and Hans Osterman Foundation for Medical Research, the Foundation for Geriatric Diseases, the Gun and Bertil Stohnes Foundation, the Gamla Tjänarinnor Foundation, the Swedish Council for Working Life and Social Research (2013-2292), and the Swedish Research Council (521-2013-8689 and 2015-03255).

Bibliography

1. Blackburn, EH, Epel, ES, and Lin, J. Human telomere biology: a contributory and interactive factor in aging, disease risks, and protection. Science 2015;350:1193–1198.
2. Okuda, K, Bardeguez, A, Gardner, JP, Rodriguez, P, Ganesh, V, Kimura, M, Skurnick, J, Awad, G, and Aviv, A. Telomere length in the newborn. Pediatric research 2002;52:377.
3. Factor-Litvak, P, Susser, E, Kezios, K, McKeague, I, Kark, JD, Hoffman, M, Kimura, M, Wapner, R, and Aviv, A. Leukocyte telomere length in newborns: implications for the role of telomeres in human disease. Pediatrics 2016:peds–2015.
4. Mitchell, C, Hobcraft, J, McLanahan, SS, Siegel, SR, Berg, A, Brooks-Gunn, J, Garfinkel, I, and Notterman, D. Social disadvantage, genetic sensitivity, and children's telomere length. Proceedings of the National Academy of Sciences 2014;111:5944–5949.
5. Gardner, M, Bann, D, Wiley, L, Cooper, R, Hardy, R, Nitsch, D, Martin-Ruiz, C, Shiels, P, Sayer, AA, Barbieri, M, et al. Gender and telomere length: systematic review and meta-analysis. Experimental gerontology 2014;51:15–27.
6. Benetos, A, Dalgård, C, Labat, C, Kark, JD, Verhulst, S, Christensen, K, Kimura, M, Horvath, K, Kyvik, KO, and Aviv, A. Sex difference in leukocyte telomere length is ablated in opposite-sex co-twins. International journal of epidemiology 2014;43:1799–1805.
7. Rewak, M, Buka, S, Prescott, J, De Vivo, I, Loucks, EB, Kawachi, I, Non, AL, and Kubzansky, LD. Race-related health disparities and biological aging: does rate of telomere shortening differ across blacks and whites? Biological psychology 2014;99:92–99.
8. Daniali, L, Benetos, A, Susser, E, Kark, JD, Labat, C, Kimura, M, Desai, KK, Granick, M, and Aviv, A. Telomeres shorten at equivalent rates in somatic tissues of adults. Nature communications 2013;4:1597.

9. Kimura, M, Stone, RC, Hunt, SC, Skurnick, J, Lu, X, Cao, X, Harley, CB, and Aviv, A. Measurement of telomere length by the Southern blot analysis of terminal restriction fragment lengths. Nature protocols 2010;5:1596.

10. Cawthon, RM. Telomere measurement by quantitative PCR. Nucleic acids research 2002;30:e47–e47.

11. O'Callaghan, NJ and Fenech, M. A quantitative PCR method for measuring absolute telomere length. Biological procedures online 2011;13:3.

12. Rufer, N, Dragowska, W, Thornbury, G, Roosnek, E, and Lansdorp, PM. Telomere length dynamics in human lymphocyte subpopulations measured by flow cytometry. Nature biotechnology 1998;16:743.

13. Cheung, I, Schertzer, M, Baross, A, Rose, AM, Lansdorp, PM, and Baird, DM. Strain-specific telomere length revealed by single telomere length analysis in Caenorhabditis elegans. Nucleic acids research 2004;32:3383–3391.

14. Honig, LS, Kang, MS, Cheng, R, Eckfeldt, JH, Thyagarajan, B, Leiendecker-Foster, C, Province, MA, Sanders, JL, Perls, T, Christensen, K, et al. Heritability of telomere length in a study of long-lived families. Neurobiology of aging 2015;36:2785–2790.

15. Broer, L, Codd, V, Nyholt, DR, Deelen, J, Mangino, M, Willemsen, G, Albrecht, E, Amin, N, Beekman, M, De Geus, EJ, et al. Meta-analysis of telomere length in 19 713 subjects reveals high heritability, stronger maternal inheritance and a paternal age effect. European Journal of Human Genetics 2013;21:1163.

16. Bakaysa, SL, Mucci, LA, Slagboom, PE, Boomsma, DI, McClearn, GE, Johansson, B, and Pedersen, NL. Telomere length predicts survival independent of genetic influences. Aging cell 2007;6:769–774.

17. Hjelmborg, JB, Dalgård, C, Möller, S, Steenstrup, T, Kimura, M, Christensen, K, Kyvik, KO, and Aviv, A. The heritability of leucocyte telomere length dynamics. Journal of medical genetics 2015;52:297–302.

18. Mangino, M, Richards, JB, Soranzo, N, Zhai, G, Aviv, A, Valdes, AM, Samani, NJ, Deloukas, P, and Spector, TD. A genome-wide association study identifies a novel locus on chromosome 18q12. 2 influencing white cell telomere length. Journal of medical genetics 2009.

19. Codd, V, Mangino, M, Harst, P van der, Braund, PS, Kaiser, M, Beveridge, AJ, Rafelt, S, Moore, J, Nelson, C, Soranzo, N, et al. Common variants near TERC are associated with mean telomere length. Nature genetics 2010;42:197.

20. Levy, D, Neuhausen, SL, Hunt, SC, Kimura, M, Hwang, SJ, Chen, W, Bis, JC, Fitzpatrick, AL, Smith, E, Johnson, AD, et al. Genome-wide association identifies OBFC1 as a locus involved in human leukocyte telomere biology. Proceedings of the National Academy of Sciences 2010;107:9293–9298.

21. Prescott, J, Kraft, P, Chasman, DI, Savage, SA, Mirabello, L, Berndt, SI, Weissfeld, JL, Han, J, Hayes, RB, Chanock, SJ, et al. Genome-wide association study of relative telomere length. PloS one 2011;6:e19635.

22. Codd, V, Nelson, CP, Albrecht, E, Mangino, M, Deelen, J, Buxton, JL, Hottenga, JJ, Fischer, K, Esko, T, Surakka, I, et al. Identification of seven loci affecting mean telomere length and their association with disease. Nature genetics 2013;45:422.

23. Pooley, KA, Bojesen, SE, Weischer, M, Nielsen, SF, Thompson, D, Amin Al Olama, A, Michailidou, K, Tyrer, JP, Benlloch, S, Brown, J, et al. A genome-wide association scan (GWAS) for mean telomere length within the COGS project: identified loci show little association with hormone-related cancer risk. Human molecular genetics 2013;22:5056–5064.

24. Cassidy, A, De Vivo, I, Liu, Y, Han, J, Prescott, J, Hunter, DJ, and Rimm, EB. Associations between diet, lifestyle factors, and telomere length in women–. The American journal of clinical nutrition 2010;91:1273–1280.

25. Latifovic, L, Peacock, SD, Massey, TE, and King, WD. The influence of alcohol consumption, cigarette smoking, and physical activity on leukocyte telomere length. Cancer Epidemiology and Prevention Biomarkers 2016;25:374–380.

26. Huzen, J, Wong, L, Veldhuisen, D, Samani, N, Zwinderman, A, Codd, V, Cawthon, R, Benus, G, Horst, I, Navis, G, et al. Telomere length loss due to smoking and metabolic traits. Journal of internal medicine 2014;275:155–163.

27. Müezzinler, A, Mons, U, Dieffenbach, AK, Butterbach, K, Saum, KU, Schick, M, Stammer, H, Boukamp, P, Holleczek, B, Stegmaier, C, et al. Smoking habits and leukocyte telomere length dynamics among older adults: results from the ESTHER cohort. Experimental gerontology 2015;70:18–25.

28. Weischer, M, Bojesen, SE, and Nordestgaard, BG. Telomere shortening unrelated to smoking, body weight, physical activity, and alcohol intake: 4,576 general population individuals with repeat measurements 10 years apart. PLoS genetics 2014;10:e1004191.

29. Strandberg, TE, Strandberg, AY, Saijonmaa, O, Tilvis, RS, Pitkälä, KH, and Fyhrquist, F. Association between alcohol consumption in healthy midlife and telomere length in older men. The Helsinki Businessmen Study. European journal of epidemiology 2012;27:815–822.

30. Saßenroth, D, Meyer, A, Salewsky, B, Kroh, M, Norman, K, Steinhagen-Thiessen, E, and Demuth, I. Sports and exercise at different ages and leukocyte telomere length in later life–Data from the Berlin Aging Study II (BASE-II). PloS one 2015;10:e0142131.

31. Nettleton, JA, Diez-Roux, A, Jenny, NS, Fitzpatrick, AL, and Jacobs Jr, DR. Dietary patterns, food groups, and telomere length in the Multi-Ethnic Study of Atherosclerosis (MESA)–. The American journal of clinical nutrition 2008;88:1405–1412.

32. Kark, JD, Goldberger, N, Kimura, M, Sinnreich, R, and Aviv, A. Energy intake and leukocyte telomere length in young adults–. The American journal of clinical nutrition 2012;95:479–487.

33. Garcìa-Calzón, S, Moleres, A, Martìnez-González, MA, Martìnez, JA, Zalba, G, and Marti, A. Dietary total antioxidant capacity is associated with leukocyte telomere length in a children and adolescent population. Clinical Nutrition 2015;34:694–699.

34. Chan, R, Woo, J, Suen, E, Leung, J, and Tang, N. Chinese tea consumption is associated with longer telomere length in elderly Chinese men. British journal of nutrition 2010;103:107–113.

35. Barden, A, O'Callaghan, N, Burke, V, Mas, E, Beilin, LJ, Fenech, M, Irish, AB, Watts, GF, Puddey, IB, Huang, RC, et al. n-3 fatty acid supplementation and leukocyte telomere length in patients with chronic kidney disease. Nutrients 2016;8:175.

36. Farzaneh-Far, R, Lin, J, Epel, ES, Harris, WS, Blackburn, EH, and Whooley, MA. Association of marine omega-3 fatty acid levels with telomeric aging in patients with coronary heart disease. Jama 2010;303:250–257.

37. Kiecolt-Glaser, JK, Epel, ES, Belury, MA, Andridge, R, Lin, J, Glaser, R, Malarkey, WB, Hwang, BS, and Blackburn, E. Omega-3 fatty acids, oxidative stress, and leukocyte telomere length: a randomized controlled trial. Brain, behavior, and immunity 2013;28:16–24.

38. Liu, JJ, Prescott, J, Giovannucci, E, Hankinson, SE, Rosner, B, Han, J, and De Vivo, I. Plasma vitamin D biomarkers and leukocyte telomere length. American journal of epidemiology 2013;177:1411–1417.

39. Crous-Bou, M, Fung, TT, Prescott, J, Julin, B, Du, M, Sun, Q, Rexrode, KM, Hu, FB, and De Vivo, I. Mediterranean diet and telomere length in Nurses' Health Study: population based cohort study. Bmj 2014;349:g6674.

40. Gu, Y, Honig, LS, Schupf, N, Lee, JH, Luchsinger, JA, Stern, Y, and Scarmeas, N. Mediterranean diet and leukocyte telomere length in a multi-ethnic elderly population. Age 2015;37:24.

41. Freitas-Simoes, TM, Ros, E, and Sala-Vila, A. Nutrients, foods, dietary patterns and telomere length: Update of epidemiological studies and randomized trials. Metabolism-Clinical and Experimental 2016;65:406–415.

42. Jackowska, M, Hamer, M, Carvalho, LA, Erusalimsky, JD, Butcher, L, and Steptoe, A. Short sleep duration is associated with shorter telomere length in healthy men: findings from the Whitehall II cohort study. PLoS One 2012;7:e47292.

43. Liang, G, Schernhammer, E, Qi, L, Gao, X, De Vivo, I, and Han, J. Associations between rotating night shifts, sleep duration, and telomere length in women. PloS one 2011;6:e23462.

44. Lee, KA, Gay, C, Humphreys, J, Portillo, CJ, Pullinger, CR, and Aouizerat, BE. Telomere length is associated with sleep duration but not sleep quality in adults with human immunodeficiency virus. Sleep 2014;37:157–166.

45. Prather, AA, Gurfein, B, Moran, P, Daubenmier, J, Acree, M, Bacchetti, P, Sinclair, E, Lin, J, Blackburn, E, Hecht, FM, et al. Tired telomeres: Poor global sleep quality, perceived stress, and telomere length in immune cell subsets in obese men and women. Brain, behavior, and immunity 2015;47:155–162.

46. Savolainen, K, Eriksson, JG, Kajantie, E, Lahti, M, and Räikkönen, K. The history of sleep apnea is associated with shorter leukocyte telomere length: the Helsinki Birth Cohort Study. Sleep medicine 2014;15:209–212.

47. Carroll, JE, Esquivel, S, Goldberg, A, Seeman, TE, Effros, RB, Dock, J, Olmstead, R, Breen, EC, and Irwin, MR. Insomnia and telomere length in older adults. Sleep 2016;39:559–564.

48. Haycock, PC, Heydon, EE, Kaptoge, S, Butterworth, AS, Thompson, A, and Willeit, P. Leucocyte telomere length and risk of cardiovascular disease: systematic review and meta-analysis. Bmj 2014;349:g4227.

49. D'mello, MJ, Ross, SA, Briel, M, Anand, SS, Gerstein, H, and Paré, G. The association between shortened leukocyte telomere length and cardio-metabolic outcomes: a systematic review and meta-analysis. Circulation: Genomic and Precision Medicine 2014:CIRCGENETICS–113.

50. Zhan, Y, Karlsson, IK, Karlsson, R, Tillander, A, Reynolds, CA, Pedersen, NL, and Hägg, S. Exploring the Causal Pathway From Telomere Length to Coronary Heart Disease: A Network Mendelian Randomization Study. Circulation Research 2017;121:214–219.

51. Haycock, PC, Burgess, S, Nounu, A, Zheng, J, Okoli, GN, Bowden, J, Wade, KH, Timpson, NJ, Evans, DM, Willeit, P, et al. Association between telomere length and risk of cancer and non-neoplastic diseases: a Mendelian randomization study. JAMA oncology 2017;3:636–651.

52. Nai-chieh, Y, Chen, BH, Song, Y, Lu, X, Chen, Y, Manson, JE, Kang, M, Howard, BV, Margolis, KL, Curb, JD, et al. A prospective study of leukocyte telomere length and risk of type 2 diabetes in postmenopausal women. Diabetes 2012;61:2998–3004.

53. Honig, LS, Kang, MS, Schupf, N, Lee, JH, and Mayeux, R. Association of shorter leukocyte telomere repeat length with dementia and mortality. Archives of neurology 2012;69:1332–1339.

54. Martin-Ruiz, C, Dickinson, HO, Keys, B, Rowan, E, Kenny, RA, and Von Zglinicki, T. Telomere length predicts poststroke mortality, dementia, and cognitive decline. Annals of neurology 2006;60:174–180.

55. Zhan, Y, Song, C, Karlsson, R, Tillander, A, Reynolds, CA, Pedersen, NL, and Hägg, S. Telomere length shortening and Alzheimer disease—a Mendelian randomization study. JAMA Neurology 2015;72:1202–1203.

56. Ma, SL, Lau, ES, Suen, EW, Lam, LCW, Leung, PC, Woo, J, and Tang, NL. Telomere length and cognitive function in southern Chinese community-dwelling male elders. Age and ageing 2013;42:450–455.

57. Yaffe, K, Lindquist, K, Kluse, M, Cawthon, R, Harris, T, Hsueh, WC, Simonsick, EM, Kuller, L, Li, R, Ayonayon, HN, et al. Telomere length and cognitive function in community-dwelling elders: findings from the Health ABC Study. Neurobiology of aging 2011;32:2055–2060.

58. Hägg, S, Zhan, Y, Karlsson, R, Gerritsen, L, Ploner, A, Lee, S van der, Broer, L, Deelen, J, Marioni, RE, Wong, A, et al. Short telomere length is associated with impaired cognitive performance in European ancestry cohorts. Translational Psychiatry 2017;7:e1100.

59. Harris, SE, Marioni, RE, Martin-Ruiz, C, Pattie, A, Gow, AJ, Cox, SR, Corley, J, Zglinicki, T von, Starr, JM, and Deary, IJ. Longitudinal telomere length shortening and cognitive and physical decline in later life: The Lothian Birth Cohorts 1936 and 1921. Mechanisms of ageing and development 2016;154:43–48.

60. Guan, JZ, Maeda, T, Sugano, M, Oyama, Ji, Higuchi, Y, Suzuki, T, and Makino, N. A percentage analysis of the telomere length in Parkinson's disease patients. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences 2008;63:467–473.

61. Wang, H, Chen, H, Gao, X, McGrath, M, Deer, D, De Vivo, I, Schwarzschild, MA, and Ascherio, A. Telomere length and risk of Parkinson's disease. Movement Disorders 2008;23:302–305.

62. Forero, DA, González-Giraldo, Y, López-Quintero, C, Castro-Vega, LJ, Barreto, GE, and Perry, G. Telomere length in Parkinson's disease: a meta-analysis. Experimental gerontology 2016;75:53–55.

63. Guan, JZ, Guan, WP, Maeda, T, Guoqing, X, GuangZhi, W, and Makino, N. Patients with multiple sclerosis show increased oxidative stress markers and somatic telomere length shortening. Molecular and cellular biochemistry 2015;400:183–187.

64. Ren, H, Collins, V, Fernandez, F, Quinlan, S, Griffiths, L, and Choo, K. Shorter telomere length in peripheral blood cells associated with migraine in women. Headache: The Journal of Head and Face Pain 2010;50:965–972.

65. Nieratschker, V, Lahtinen, J, Meier, S, Strohmaier, J, Frank, J, Heinrich, A, Breuer, R, Witt, SH, Nöthen, MM, Rietschel, M, et al. Longer telomere length in patients with schizophrenia. Schizophrenia research 2013;149:116–120.

66. Kota, LN, Purushottam, M, Moily, NS, and Jain, S. Shortened telomere in unremitted schizophrenia. Psychiatry and clinical neurosciences 2015;69:292–297.

67. Fernandez-Egea, E, Bernardo, M, Heaphy, CM, Griffith, JK, Parellada, E, Esmatjes, E, Conget, I, Nguyen, L, George, V, Stöppler, H, et al. Telomere length and pulse pressure in newly diagnosed, antipsychotic-naive patients with nonaffective psychosis. Schizophrenia bulletin 2009;35:437–442.

68. Ridout, KK, Ridout, SJ, Price, LH, Sen, S, and Tyrka, AR. Depression and telomere length: a meta-analysis. Journal of affective disorders 2016;191:237–247.

69. Schutte, NS and Malouff, JM. The association between depression AND leukocyte telomere length: a meta-analysis. Depression and anxiety 2015;32:229–238.

70. Zhang, C, Doherty, JA, Burgess, S, Hung, RJ, Lindström, S, Kraft, P, Gong, J, Amos, CI, Sellers, TA, Monteiro, AN, et al. Genetic determinants of telomere length and risk of common cancers: a Mendelian randomization study. Human molecular genetics 2015;24:5356–5366.

71. Cawthon, RM, Smith, KR, O'Brien, E, Sivatchenko, A, and Kerber, RA. Association between telomere length in blood and mortality in people aged 60 years or older. The Lancet 2003;361:393–395.

72. Martin-Ruiz, CM, Gussekloo, J, Heemst, D, Zglinicki, T, and Westendorp, RG. Telomere length in white blood cells is not associated with morbidity or mortality in the oldest old: a population-based study. Aging cell 2005;4:287–290.

73. Harris, SE, Deary, IJ, MacIntyre, A, Lamb, KJ, Radhakrishnan, K, Starr, JM, Whalley, LJ, and Shiels, PG. The association between telomere length, physical health, cognitive ageing, and mortality in non-demented older people. Neuroscience letters 2006;406:260–264.

74. Houben, JM, Giltay, EJ, Rius-Ottenheim, N, Hageman, GJ, and Kromhout, D. Telomere length and mortality in elderly men: the Zutphen Elderly Study. Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences 2011;66:38–44.

75. Epel, ES, Merkin, SS, Cawthon, R, Blackburn, EH, Adler, NE, Pletcher, MJ, and Seeman, TE. The rate of leukocyte telomere shortening predicts mortality from cardiovascular disease in elderly men. Aging (Albany NY) 2009;1:81.

76. Willeit, P, Willeit, J, Mayr, A, Weger, S, Oberhollenzer, F, Brandstätter, A, Kronenberg, F, and Kiechl, S. Telomere length and risk of incident cancer and cancer mortality. Jama 2010;304:69–75.

77. Willeit, P, Willeit, J, Kloss-Brandstätter, A, Kronenberg, F, and Kiechl, S. Fifteen-year follow-up of association between telomere length and incident cancer and cancer mortality. Jama 2011;306:42–44.

78. Rode, L, Nordestgaard, BG, and Bojesen, SE. Peripheral blood leukocyte telomere length and mortality among 64 637 individuals from the general population. JNCI: Journal of the National Cancer Institute 2015;107.

79. Needham, BL, Rehkopf, D, Adler, N, Gregorich, S, Lin, J, Blackburn, EH, and Epel, ES. Leukocyte telomere length and mortality in the National Health and Nutrition Examination Survey, 1999–2002. Epidemiology (Cambridge, Mass.) 2015;26:528.

80. Carty, CL, Kooperberg, C, Liu, J, Herndon, M, Assimes, T, Hou, L, Kroenke, CH, LaCroix, A, Kimura, M, Aviv, A, et al. Leukocyte telomere length and risks of incident coronary heart disease and mortality in a racially diverse population of postmenopausal women. Arteriosclerosis, thrombosis, and vascular biology 2015:ATVBAHA–115.

81. Lambert, JC, Ibrahim-Verbaas, CA, Harold, D, Naj, AC, Sims, R, Bellenguez, C, Jun, G, DeStefano, AL, Bis, JC, Beecham, GW, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nature genetics 2013;45:1452.

82. Schunkert, H, König, IR, Kathiresan, S, Reilly, MP, Assimes, TL, Holm, H, Preuss, M, Stewart, AF, Barbalic, M, Gieger, C, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nature genetics 2011;43:333.

83. Dupuis, J, Langenberg, C, Prokopenko, I, Saxena, R, Soranzo, N, Jackson, AU, Wheeler, E, Glazer, NL, Bouatia-Naji, N, Gloyn, AL, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nature genetics 2010;42:105.

84. Scott, RA, Lagou, V, Welch, RP, Wheeler, E, Montasser, ME, Luan, J, Mägi, R, Strawbridge, RJ, Rehnberg, E, Gustafsson, S, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nature genetics 2012;44:991.

85. Morris, AP, Voight, BF, Teslovich, TM, Ferreira, T, Segre, AV, Steinthorsdottir, V, Strawbridge, RJ, Khan, H, Grallert, H, Mahajan, A, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nature genetics 2012;44:981.

86. Soranzo, N, Sanna, S, Wheeler, E, Gieger, C, Radke, D, Dupuis, J, Bouatia-Naji, N, Langenberg, C, Prokopenko, I, Stolerman, E, et al. Common variants at 10 genomic loci influence hemoglobin A1C levels via glycemic and nonglycemic pathways. Diabetes 2010;59:3229–3239.

87. Locke, AE, Kahali, B, Berndt, SI, Justice, AE, Pers, TH, Day, FR, Powell, C, Vedantam, S, Buchkovich, ML, Yang, J, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature 2015;518:197.

88. Willer, CJ, Schmidt, EM, Sengupta, S, Peloso, GM, Gustafsson, S, Kanoni, S, Ganna, A, Chen, J, Buchkovich, ML, Mora, S, et al. Discovery and refinement of loci associated with lipid levels. Nature genetics 2013;45:1274.

89. Horikoshi, M, Mägi, R, Bunt, M van de, Surakka, I, Sarin, AP, Mahajan, A, Marullo, L, Thorleifsson, G, Hägg, S, Hottenga, JJ, et al. Discovery and fine-mapping of glycaemic and obesity-related trait loci using high-density imputation. PLoS genetics 2015;11:e1005230.

90. Surakka, I, Horikoshi, M, Mägi, R, Sarin, AP, Mahajan, A, Lagou, V, Marullo, L, Ferreira, T, Miraglio, B, Timonen, S, et al. The impact of low-frequency and rare variants on lipid levels. Nature genetics 2015;47:589.

91. Gold, CH, Malmberg, B, McClearn, GE, Pedersen, NL, and Berg, S. Gender and health: a study of older unlike-sex twins. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences 2002;57:S168–S176.

92. Finkel, D and Pedersen, NL. Processing speed and longitudinal trajectories of change for cognitive abilities: The Swedish Adoption/Twin Study of Aging. Aging Neuropsychology and Cognition 2004;11:325–345.

93. Pedersen, NL, McClearn, GE, Plomin, R, Nesselroade, JR, Berg, S, and DeFaire, U. The Swedish adoption twin study of aging: an update. Acta Geneticae Medicae et Gemellologiae: Twin Research 1991;40:7–20.

94. McClearn, GE, Johansson, B, Berg, S, Pedersen, NL, Ahern, F, Petrill, SA, and Plomin, R. Substantial genetic influence on cognitive abilities in twins 80 or more years old. Science 1997;276:1560–1563.

95. Roberts, RO, Geda, YE, Knopman, DS, Cha, RH, Pankratz, VS, Boeve, BF, Ivnik, RJ, Tangalos, EG, Petersen, RC, and Rocca, WA. The Mayo Clinic Study of Aging: design and sampling, participation, baseline measures and sample characteristics. Neuroepidemiology 2008;30:58–69.

96. Sonnega, A, Faul, JD, Ofstedal, MB, Langa, KM, Phillips, JW, and Weir, DR. Cohort profile: The health and retirement study (HRS). International journal of epidemiology 2014;43:576–585.

97. Pedersen, NL, Plomin, R, Nesselroade, JR, and McClearn, GE. A quantitative genetic analysis of cognitive abilities during the second half of the life span. Psychological Science 1992;3:346–353.

98. Reynolds, CA, Finkel, D, McArdle, JJ, Gatz, M, Berg, S, and Pedersen, NL. Quantitative genetic analysis of latent growth curve models of cognitive abilities in adulthood. Developmental psychology 2005;41:3.

99. Mielke, MM, Hagen, CE, Wennberg, AM, Airey, DC, Savica, R, Knopman, DS, Machulda, MM, Roberts, RO, Jack, CR, Petersen, RC, et al. Association of plasma total tau level with cognitive decline and risk of mild cognitive impairment or dementia in the Mayo Clinic Study on Aging. JAMA neurology 2017;74:1073–1080.

100. Weir, DR. Health and Retirement Study Imputation of Cognitive Functioning Measures. 2017.

101. Davey Smith, G and Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? International journal of epidemiology 2003;32:1–22.

102. Sheehan, NA, Didelez, V, Burton, PR, and Tobin, MD. Mendelian randomisation and causal inference in observational epidemiology. PLoS medicine 2008;5:e177.

103. Burgess, S, Daniel, RM, Butterworth, AS, Thompson, SG, and Consortium, EI. Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways. International journal of epidemiology 2014;44:484–495.

104. Burgess, S, Scott, RA, Timpson, NJ, Smith, GD, Thompson, SG, Consortium, EI, et al. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. European journal of epidemiology 2015;30:543–552.

105. Johnson, T. Efficient calculation for multi-SNP genetic risk scores. 2012.

106. Bowden, J, Davey Smith, G, and Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. International journal of epidemiology 2015;44:512–525.

107. Bowden, J, Davey Smith, G, Haycock, PC, and Burgess, S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genetic epidemiology 2016;40:304–314.

108. Grimm, KJ, Ram, N, and Estabrook, R. Growth modeling: Structural equation and multilevel modeling approaches. Guilford Publications, 2016.

109. Liu, XR, Pawitan, Y, and Clements, M. Parametric and penalized generalized survival models. Statistical methods in medical research 2016:0962280216664760.

110. Liu, XR, Pawitan, Y, and Clements, MS. Generalized survival models for correlated time-to-event data. Statistics in medicine 2017;36:4743–4762.

111. Sjölander, A, Lichtenstein, P, Larsson, H, and Pawitan, Y. Between–within models for survival analysis. Statistics in medicine 2013;32:3067–3076.

112. Dahlqwist, E, Pawitan, Y, and Sjölander, A. Regression standardization and attributable fraction estimation with between-within frailty models for clustered survival data. Statistical methods in medical research 2017:0962280217727558.

113. Jensen, H, Brookmeyer, R, Aaby, P, and Andersen, PK. Shared frailty model for left-truncated multivariate survival data. University of Copenhagen. Department of Biostatistics, 2004.

114. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria, 2017. URL: `https://www.R-project.org/`.

115. Lumley, T. rmeta: Meta-analysis. R package version 2.16. 2012. URL: `https://CRAN.R-project.org/package=rmeta`.

116. Johnson, T. gtx: Genetics ToolboX. R package version 0.0.8. 2013. URL: `https://CRAN.R-project.org/package=gtx`.

117. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016. URL: `http://ggplot2.org`.

118. Clements, M and Liu, XR. rstpm2: Generalized Survival Models. R package version 1.4.1. 2017. URL: `https://CRAN.R-project.org/package=rstpm2`.

119. Pierce, BL and VanderWeele, TJ. The effect of non-differential measurement error on bias, precision and power in Mendelian randomization studies. International journal of epidemiology 2012;41:1383–1393.

120. Zhao, Q, Wang, J, Bowden, J, and Small, DS. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. arXiv preprint arXiv:1801.09652 2018.

121. VanderWeele, T. Explanation in causal inference: methods for mediation and interaction. Oxford University Press, 2015.